

PTO 04-335

Japanese Kokai Patent Application
No. Hei 7[1995]-271798

INFORMATION RETRIEVAL TECHNIQUE EVALUATION METHOD AND DEVICE FOR
SAME

Hideaki Ozawa and Toru Nakagawa

UNITED STATES PATENT AND TRADEMARK OFFICE
WASHINGTON, D.C. NOVEMBER 2003
TRANSLATED BY THE RALPH MCELROY TRANSLATION COMPANY

**JAPANESE PATENT OFFICE
PATENT JOURNAL (A)
KOKAI PATENT APPLICATION NO. HEI 7[1995]-271798**

Int. Cl. ⁶ :	G 06 F 17/30 G 06 F 15/403
Sequence No. for Office Use:	9194-5L
Filing No.:	Hei 6[1994]-60682
Filing Date:	March 30, 1994
Publication Date:	October 20, 1995
No. of Claims:	6 (Total of 19 pages; OL)
Examination Request:	Not filed

**INFORMATION RETRIEVAL TECHNIQUE EVALUATION METHOD AND DEVICE FOR
SAME**

[Joho kensaku shuho hyoka hoho oyobi sono sochi]

Inventors:	Hideaki Ozawa and Toru Nakagawa
Applicant:	000004226 NEC Corp.

[There are no amendments to this patent.]

Claims

/2*

1. An information retrieval technique evaluation method characterized in that it is an evaluation method for the retrieval efficiency of an information retrieval technique used on an information retrieval system that acquires retrieved data that match data to be retrieved given in advance from a database, that calculates the retrieval efficiency of respective retrieval techniques from an external evaluation value that indicates whether the data in the database match the data to be retrieved determined by at least one or more persons, and a value as to whether or not the data in

* [Numbers in right margin indicate pagination in the original foreign language text.]

the database match the data to be retrieved determined by the information retrieval technique, and that determines the merits of the retrieval technique,

the number of external values where it is determined by at least one or more persons that the data to be retrieved and any one datum in the database match, and the number of external values where it is determined that they do not match are each counted,

if the judgment as to whether or not they agree with the data to be retrieved that is determined by the information retrieval technique for the one datum is the same as the judgments on the majority side as to whether or not they agree with the retrieval conditions produced by a given external evaluation value, a constant value given in advance is used as the data evaluation value, and if the same as the minority side, a value obtained with a formula given in advance from the number of external evaluation values on the majority and minority sides is used as a one datum evaluation value,

the aforementioned one datum evaluation value is created for the combination of all the given data to be retrieved and the data in the database and the average value of all the one datum evaluation values is calculated,

retrieval efficiency is expressed using one index by using said average value as the retrieval efficiency evaluation value for the information retrieval technique,

and the merits of each information retrieval technique are determined by sequencing the evaluation value calculated for at least one or more information retrieval techniques.

2. The information retrieval technique evaluation method mentioned in Claim 1 characterized in that when the aforementioned one datum evaluation value is calculated, the actual evaluation value is calculated from the number of people on the minority side and the number of people on the majority side in accordance with a predetermined formula,

a predetermined number of people is subtracted from the number of people on the majority side, the value subtracted from the majority side is added to the number of people on the minority side, and an imaginary evaluation value is calculated in accordance with a predetermined formula,

the difference between the imaginary evaluation value and the actual evaluation value is used as the degree of wobble of the one datum evaluation value and the average value of the degree of wobble for the combination of all retrieval conditions and the data in the database is calculated,

the retrieval efficiency with a confidence interval that increases or decreases the value of the extent of wobble is calculated for the aforementioned retrieval efficiency value,

and the merits of the retrieval efficiency of multiple information retrieval techniques are judged by calculating the maximum value and the minimum value for the retrieval efficiency with said confidence interval and one retrieval efficiency value for each retrieval technique with a formula given in advance from the aforementioned data evaluation values.

3. The information retrieval technique evaluation method mentioned in Claim 1 or 2 characterized in that the corresponding number of people that agrees with the data to be retrieved and the corresponding number of people that do not agree are compared by the external evaluation value in the calculation of the aforementioned one datum evaluation value, and when they are the same number, a predetermined value is used as the one datum evaluation value to calculate the retrieval efficiency of each retrieval technique.

4. An information retrieval technique device characterized in that it is an information retrieval technique evaluation device which is a retrieval efficiency evaluation device for an information retrieval technique used on an information retrieval system that acquires retrieved data that matches data to be retrieved given in advance, that calculates the retrieval efficiency of respective retrieval techniques from an external evaluation value that indicates whether the data in the database match the data to be retrieved determined by at least one or more persons, and a value as to whether or not the data in the database match the data to be retrieved determined by the information retrieval technique, and that determines the merits of the retrieval technique,

and that is furnished with: an input means for inputting the result of the judgment as to whether or not something should be retrieved from at least one or more persons,

a counting means that counts the number of people that think something should be retrieved and the number of people that think something should not be retrieved from said means for input of human judgment result between individual retrieval results and individual data to be retrieved,

a determination means that compares the number of people that think something should be retrieved counted by said counting means and the number of people that think it should not be retrieved, and that determines that said data to be retrieved should be retrieved when the number of people that think it should be retrieved is at or above a predetermined ratio,

a judgment means that judges whether the judgment that the information retrieval technique should retrieve for the aforementioned retrieval conditions and the data to be retrieved agrees or does not agree,

a constant value generation means that generates a predetermined constant when the result of said judgment is that there is agreement,

a penalty value generation means that generates a penalty value according to a predetermined formula when the result of judgment by the aforementioned judgment means is that there is not agreement,

an average value calculation means that finds the average value of the constant value and the penalty value obtained for the combination of all given retrieval conditions and data to be retrieved as an average value,

and a sorting means that places the average values obtained for each retrieval technique in a prescribed order.

5. The information retrieval technique evaluation device mentioned in Claim 4 characterized in that it is furnished with: a calculation means that determines the magnitude in the difference of human judgment according to a predetermined formula after the number of people that think that the results of human judgment should be retrieved and the number of people that think they should not be retrieved are counted,

a comparison means that compares the number of people that think something should be retrieved and the number of people that think it should not be retrieved in order to provide noise to the human judgment,

an addition and subtraction means that subtracts a predetermined constant value from the value of the number of people on the majority side and adds a predetermined constant to the number of people on the minority side as a result of comparing the number of people by said comparison means,

a recalculation means that again calculates the magnitude of the difference in judgment between humans according to a predetermined formula from the number of people that think it should be retrieved that is reset by said addition and subtraction means and the number of people that think it should not be retrieved,

a second average value calculation means that finds the average value of the result of calculating the magnitude of the difference between the human judgment produced by the aforementioned recalculation means and the judgment containing noise for the combination of all retrieval conditions and retrieval data,

a minimum and maximum value calculation means that calculates the minimum value and maximum value of the average value with a predetermined formula from the average value obtained by the aforementioned average value calculation means and the average value obtained by the aforementioned second average value calculation means,

an order calculation means that calculates the respective order of each retrieval technique by arranging the average value, minimum value, and maximum value in order from the smallest,

and an order determination means that uses the order value of each retrieval technique for each of the three values of average value, minimum value and maximum value as one value with a predetermined formula.

6. The information retrieval technique evaluation device mentioned in Claim 4 or 5 characterized in that it is furnished with: a comparison means that compares the number of people that think it should be retrieved and the number of people that think it should not be retrieved between one datum to be retrieved and one [more] datum to be retrieved,

and a constant generation means that generates a predetermined value if the result of comparison by said comparison means is the same number.

Detailed explanation of the invention

[0001]

Industrial application field

This invention relates to an information retrieval technique evaluation method and device for same for automatically selecting the optimal information retrieval technique when multiple information retrieval techniques are available. In particular, it relates to an evaluation method and device for a database retrieval technique where the relationship between retrieval conditions and data that should be retrieved changes often according to [the needs of the] user, as represented by a current events information database, such as for newspaper articles.

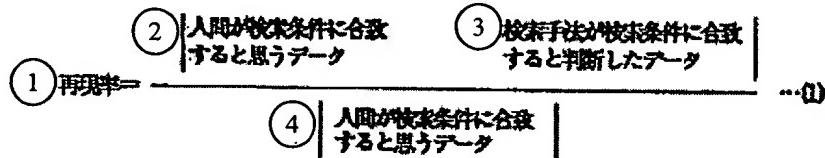
[0002]

Prior art

In the past, in evaluating information retrieval techniques, two indices, called the recall ratio represented by Formula (1) below and the relevance factor represented by Formula (2), would have been judged in combination. (There are many, for example

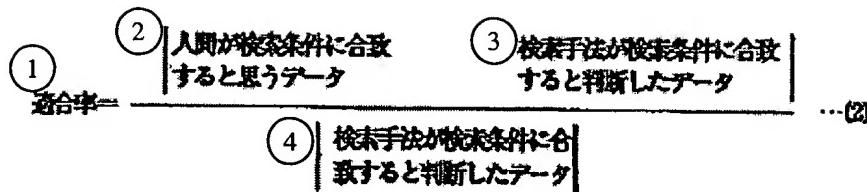
Gerard Salton (Ed.). *The SMART Retrieval System - Experiments in Automatic Document Processing*, Englewood Cliffs, NJ, Prentice-Hall, 1971. B.Masand, G.Linoff, D.Waltz. *Classifying News Stories using Memory Based Reasoning*. Proceedings, 15th Int'l SIGIR 1992, pp.59-56, 1992

Formula (1)



- Key:
- 1 Recall ratio
 - 2 Data that people think matches retrieved data
 - 3 Data judged to match the retrieved data by the retrieval technique
 - 4 Data that people think matches the retrieval conditions

Formula (2)



- Key:
- 1 Relevance factor
 - 2 Data that people think matches the retrieval conditions
 - 3 Data judged to match the retrieval conditions by the retrieval technique
 - 4 Data judged to match the retrieval conditions by the retrieval technique

[0003]

Problems to be solved by the invention

However, with indices produced by the recall ratio and the relevance factor, the numerator values are the same and the difference in denominator value only affects the indices. The numerator value with the recall ratio is the number of data that should be retrieved based on human judgment, and in the case of the relevance factor, it is the number of data that should be retrieved in the judgment of the retrieval technique. When the retrieval technique conditions are relaxed to increase the recall ratio without changing human judgment, many results will be retrieved, so the relevance factor drops. Conversely, when the relevance factor is increased, the retrieval technique conditions will become more strict, so there is a tradeoff relationship in recall ratio and relevance factor. For this reason, when the merits and disadvantages of various retrieval techniques are evaluated, which technique is suitable could not be judged by whether a technique with higher relevance factor or a technique with higher recall ratio should be used.

[0004]

In particular, the relevance factor uses a “value produced by the retrieval technique” for the denominator. So for example, even if the same value of 30% is used, in a case where 9 are correct in the judgment of humans, 30 data are obtained as the retrieved result and all 9 that humans think should be retrieved are included, and a case where only 10 data are retrieved and 3 correct data are included, it is not possible to judge which case represents the better retrieval technique.

[0005]

In the recall ratio, too, even with retrieval from the same retrieved data by a person who is retrieving, such as for current events information, when the content of the retrieval results they

want are different, the denominator of the recall factor changes, and the value of the recall factor changes, which would be a problem.

[0006]

This problem, when the retrieval efficiency evaluation method is represented as in Figure 2, is caused by the fact that the data that a person thinks should be retrieved only receives attention when it should be retrieved in response to the information retrieval technique, and the fact that the human judgment shifts and data that should not be retrieved are not retrieved is not included in the evaluation technique.

[0007]

The purpose of this invention, in consideration of the aforementioned problems, is to provide an information retrieval technique evaluation method and device for same that can accurately judge the optimal information retrieval technique.

[0008]

Means to solve the problems

/4

With this invention, with Claim 1 for achieving the aforementioned purpose, an information retrieval technique evaluation method is proposed that is an evaluation method for the retrieval efficiency of an information retrieval technique used on an information retrieval system that acquires retrieved data that match data to be retrieved given in advance from a database, that calculates the retrieval efficiency of respective retrieval techniques from an external evaluation value that indicates whether the data in the database match the data to be retrieved determined by at least one or more persons, and a value as to whether or not the data in the database match the data to be retrieved determined by the information retrieval technique, and that determines the merits of the retrieval technique, the number of external values where it is determined by at least one or more persons that the data to be retrieved and any one datum in the database match, and the number of external values where it is determined that they do not match are each counted, if the judgment as to whether or not the data to be retrieved that is determined by the information retrieval technique for the one datum is matched is the same as the judgments on the majority side as to whether or not retrieval conditions produced by a given external evaluation value are matched, a constant value given in advance is used as the data evaluation value. If it is the same as the minority side, a value obtained with a formula given in advance from the number of external evaluation values on the majority and minority sides is used as a one datum evaluation value. The aforementioned one datum evaluation value is created for the combination of all the given data to be retrieved and the data in the database and the average value of all the one datum evaluation

values is calculated. Retrieval efficiency is expressed using one index by using said average value as the retrieval efficiency evaluation value for the information retrieval technique, and the merits of each information retrieval technique are determined by sequencing the evaluation value calculated for at least one or more information retrieval techniques.

[0009]

And with Claim 2, in the information retrieval technique evaluation method mentioned in Claim 1, an information retrieval technique evaluation method is proposed such that when the aforementioned data evaluation value is calculated, the actual evaluation value is calculated from the number of people on the minority side and the number of people on the majority side of the external evaluation value according to a predetermined formula, a predetermined number of people is subtracted from the number of people on the majority side, the value subtracted from the majority side is added to the number of people on the minority side, and an imaginary evaluation value is calculated according to a predetermined formula. The difference between the imaginary evaluation value and the actual evaluation value is used as the extent of wobble of the one datum evaluation value. The average value of the extent of wobble for the combination of all the retrieval conditions and the data in the database is calculated and retrieval efficiency with a confidence interval where the extend of wobble in the value of the aforementioned retrieval efficiency value is increased and decreased is calculated. The merits of retrieval efficiency of multiple information retrieval techniques are judged by calculating one retrieval efficiency value for each retrieval technique with a predetermined formula from the maximum value and minimum value of said retrieval efficiency with confidence interval and the aforementioned data evaluation value.

[0010]

And with Claim 3, in the information retrieval technique evaluation method mentioned in Claim 1 or 2, an information retrieval technique evaluation method is proposed where, in the calculation of the aforementioned data evaluation value, the number of corresponding people that agree with the retrieval data and the corresponding number of people that do not agree are compared by the external evaluation value, and when they are the same, the retrieval efficiency of each retrieval technique is calculated with a predetermined value as a one datum evaluation value.

[0011]

Also, with Claim 4, an information retrieval technique evaluation device is proposed using an information retrieval technique used on an information retrieval system that acquires retrieved data that match data to be retrieved given in advance, and is an information retrieval technique evaluation device that calculates the retrieval efficiency of respective retrieval techniques from an

external evaluation value that indicates whether the data in the database match the data to be retrieved determined by at least one or more persons, and a value as to whether or not the data in the database match the data to be retrieved determined by the information retrieval technique, and that determines the merits of the retrieval technique. There are furnished: an input means for inputting the result of the judgment as to whether or not something should be retrieved from at least one or more persons, a counting means that counts the number of people that think it should be retrieved and the number of people that think it should not be retrieved from said means for input of human judgment result between individual retrieval results and individual data to be retrieved, a determination means that compares the number of people that think something should be retrieved counted by said counting means and the number of people that think it should not be retrieved, and that determines that said data to be retrieved should be retrieved when the number of people that think it should be retrieved is at or above a predetermined ratio, a judgment means that judges whether the judgment that the information retrieval technique should retrieve for the aforementioned retrieval conditions and the data to be retrieved agrees or does not agree, a constant value generation means that generates a predetermined constant when the result of said judgment is that there is agreement, a penalty value generation means that generates a penalty value according to a predetermined formula when the result of judgment by the aforementioned judgment means is that there is not agreement, an average value calculation means that finds the average value of the constant value and the penalty value obtained for the combination of all given retrieval conditions and data to be retrieved as an average value, and a sorting means that places the average values obtained for each retrieval technique in a prescribed order.

[0012]

Also, with Claim 5, an information retrieval technique evaluation device is proposed such that, in the information retrieval technique evaluation device mentioned in Claim 4, there are furnished: a calculation means that calculates the magnitude in the difference of human judgment according to a predetermined formula after the number of people that think that the results of human judgment should be retrieved and the number of people that think they should not be retrieved are counted, a comparison means that compares the number of people that think it should be retrieved and the number of people that think it should not be retrieved in order to provide noise to the human judgment, an addition and subtraction means that subtracts a predetermined constant value from the value of the number of people on the majority side and adds a predetermined constant to the number of people on the minority side as a result of comparing the number of people by said comparison means, a recalculation means that again calculates the magnitude of the difference in judgment between humans according to a predetermined formula from the number of people that think it should be retrieved that is reset by said addition and subtraction means and the

number of people that think it should not be retrieved, a second average value calculation means that finds the average value of the result of calculating the magnitude of the difference between the human judgment produced by the aforementioned recalculation means and the judgment containing noise for the combination of all retrieval conditions and retrieval data, a minimum and maximum value calculation means that calculates the minimum value and maximum value of the average value with a predetermined formula from the average value obtained by the aforementioned average value calculation means and the average value obtained by the aforementioned second average value calculation means, an order calculation means that calculates the respective order of each retrieval technique by arranging the average value, minimum value, and maximum value in order from the smallest, and an order determination means that uses the order value of each retrieval technique for each of the three values of average value, minimum value and maximum value as one value with a predetermined formula.

[0013]

In addition, with Claim 6, an information retrieval technique evaluation device is proposed where, in the information retrieval technique evaluation device mentioned in Claim 4 or 5, a comparison means is furnished that compares the number of people that think it should be retrieved and the number of people that think it should not be retrieved, between one datum to be retrieved and one [more] datum to be retrieved, as well as a constant generation means that generates a predetermined value if the comparison results of said comparison means are the same number.

[0014]

Operation

With Claim 1 of this invention, the number of external evaluation values where it is determined by at least one or more persons that data to be retrieved and any one datum in a database match the data to be retrieved, and the number of external evaluation values where it is determined that they do not match are each counted. If the judgment as to whether or not there is agreement with the data to be retrieved determined by the information retrieval technique for the one datum is the same as the majority side of judgments as to whether or not there is agreement with the retrieval conditions based on the given external evaluation values, a constant value given in advance is used as the one datum evaluation value. If it is the same as the minority side, a value obtained with a formula given in advance from the number of external evaluation values on the majority side and the minority side is used as the data evaluation value, and the aforementioned one datum evaluation value is created for the combination of all given data to be retrieved and the data in the database. In addition, the average value of all the data evaluation values is calculated,

and by using this as the retrieval efficiency evaluation value for the information retrieval technique, the retrieval efficiency is represented using one index. By sequencing the valuation values calculated for at least one or more information retrieval techniques, the merits of each type of information retrieval technique are determined.

[0015]

Also, with Claim 2, when the aforementioned one datum evaluation value is calculated, an actual evaluation value is calculated from the value of the number of people on the minority side and the number of people on the majority side of the external evaluation value according to a predetermined formula. A predetermined number of people is subtracted from the number of people on the majority side, the value subtracted from the majority side is added to the number of people on the minority side, and an imaginary evaluation value according to a predetermined formula is calculated. The difference between said imaginary evaluation value and the actual evaluation value is used as the extent of wobble of the one datum evaluation value. The average value of the extent of wobble for the combination of all retrieval conditions and the data in the database is calculated, the retrieval efficiency with a confidence interval where the value of the extent of wobble for the aforementioned retrieval efficiency value is further increased or decreased is calculated, and one retrieval efficiency value is calculated for each retrieval technique with a predetermined formula from the maximum value and minimum value of said retrieval efficiency with confidence interval and from the aforementioned one datum evaluation value so that the merits of retrieval efficiency of multiple information retrieval techniques are judged.

[0016]

Also, with Claim 3, in the calculation of the aforementioned data evaluation value, the corresponding number of people that agree with the data to be retrieved and the corresponding number of people that do not agree is compared by the external evaluation value. When they are the same number, a predetermined value is used as the one datum evaluation value and the retrieval efficiency of each retrieval technique is calculated.

[0017]

Also, with Claim 4 whether said data to be retrieved should be retrieved or not retrieved is determined by the counting means from data for the judgment of at least one person as to whether or not something input by an input means should be retrieved. So for judgments as to whether not something should be retrieved as judged by a human, the number of people for each is counted, a predetermined value for judging whether or not it should be retrieved and the measured value of the ratio of the number of people that think it should or should not be retrieved are compared by a

determination means, and it is determined whether said data to be retrieved are data that should be retrieved. Next, the judgment result as to whether or not the input information retrieval system should be performed and the judgment result as to whether or not something should be retrieved determined from human judgment are compared by the judgment means, and it is determined whether the judgment result obtained by the retrieval system is in agreement with the judgment produced by humans. In addition, when the human judgment and the judgment by the retrieval technique agree, a predetermined constant value is given by a constant value generation means. When the human and retrieval technique judgments are different, a penalty value is calculated with a predetermined formula from the value of the ratio of the various numbers of people for the judgments as to whether or not something should be retrieved in the judgment of the humans. For example, assuming that the constant value when the judgments agree is 0, the penalty value when they do not agree should be a value larger than 0. Because of this, when the judgment by the retrieval technique agrees completely with the human judgment as to whether or not something should be retrieved, 0 is always given to individual retrieval conditions and to individual data to be retrieved. For data when the judgments do not agree, a value larger than 0 is obtained, so by calculating the average value of the penalty values with an average value calculation means, the evaluation value of retrieval technique will be 0 when the judgments agree perfectly, penalty values are used that are larger for retrieval techniques that have more cases where the judgments do not agree, and the capability of each retrieval technique is expressed numerically as one numerical value. In addition, by sequencing the penalty values calculated, for example, in order of the smallest, with a sorting means, the capability of at least one or more retrieval techniques can be compared.

[0018]

Also, with Claim 5, when a judgment as to whether or not to retrieve is performed using individual retrieval conditions and individual data to be retrieved, an actual evaluation value, that is, the magnitude of the difference in judgments between people, is calculated with a predetermined formula from the ratio of the number of people that think it should be retrieved and the number of people that think it should not be retrieved by a calculation means. The number of people that think it should be retrieved and the number of people that think it should not be retrieved are compared by a comparison means. Based on said comparison results, a predetermined number of people is subtracted from the number of people on the majority side of the number of people who think it should be retrieved and the number of people who think it should not be retrieved, and a predetermined number of people is added to the minority side by an addition and subtraction means. The result is again calculated by a recalculation means in accordance with a predetermined formula from the ratio of the changed number of people, an imaginary evaluation value, that is, the

magnitude of the difference in human judgments including noise is calculated, and the results of judgments by humans that are input and the results with noise included in the judgment are created. In addition, the magnitude of the differences of these two judgments is measured with a predetermined formula, and when the human judgment is slightly different, how much the penalty value will be affected is calculated. The effect of this value on the ratio of the number of people is large when the judgment of whether or not something should be retrieved is made by a sufficiently large number of people, even when an extremely small number of people move to the majority side or the minority side, and the effect of the value on the calculated penalty value is also large. Thus, by calculating the magnitude of the difference between the human judgment with noise added and the human judgment input, to what extent confidence can be placed in it can be numerically expressed. The minimum value and maximum value of the evaluation value are calculated with a minimum and maximum value calculation means from the retrieval efficiency evaluation value and the value that indicates the degree of confidence. When the degree of confidence is low, the retrieval efficiency can be represented by data for a broad range of evaluation values, and when the degree of confidence is high, by data for a narrow range of evaluation values. In addition, with an order calculation means that sequences the efficiency values, and with an order determination means that calculates the order of retrieval efficiency as one value from the order in each value with a predetermined formula, between minimum values, maximum values, and evaluation values, with an evaluation value with a low degree of confidence, for example, the possibility that the maximum value will be very large is high. So within the order of retrieval techniques for each of the values of maximum value, minimum value, and evaluation value, it is judged that a retrieval technique with a smaller maximum value indicates better retrieval efficiency. The result is that even when there is a difference in judgments by people, the merits and disadvantage of one or more retrieval techniques can be determined automatically. Because of this, when the results produced by an information retrieval technique and human judgment are different, a penalty value is calculated from the ratio of people who think something should be retrieved and the ratio of people who think it should not be retrieved, but if the value of this penalty is viewed statistically, the reliability changes according to the number of people making the judgment. So with the device mentioned in aforementioned Claim 4, a judgment as to whether or not something should be retrieved must necessarily be made by the same number of people, but with the device mentioned in Claim 5 there is no such restriction.

[0019]

Also, with Claim 6, when the majority and minority are determined, the number of people who think that something should be retrieved and the number of people who think it should not be retrieved between one datum to be retrieved and one [more] datum to be retrieved are compared by

a comparison means. Based on the results of said comparison, when the number of these people is the same, a predetermined constant is generated by a constant generation means, so that an evaluation value is calculated even when there is no majority or minority. Because of this, even when the number of people judge that it should be retrieved and the number of people that think it should not be retrieved are the same, it is possible to judge between data that should be retrieved and data that should not be retrieved.

[0020]

Application example

Below, an application example of this invention will be explained with reference to the figures. Figure 1 is a block diagram that shows an information retrieval technique evaluation device in a first application example of this invention. Said device is constituted from a computer that primarily a CPU and from programs. In the figure, (10) is an external evaluation value and retrieval technique agreement determination module, (2) is an external evaluation value totaling module, (3) is a relational judgment module, (4) is a constant value generation module, (5) is a penalty calculation module, (6) is an average penalty calculation module, (7) is an evaluation value buffer, and (8) is a sorting module. Note that here, each module is constituted as a processing unit of a program (software), but they can also be constituted with hardware.

[0021]

External evaluation value and retrieval technique agreement determination module (1) compares the judgment as to whether or not something should be retrieved produced by the retrieval technique and the judgment given by a human.

[0022]

External evaluation value totaling module (2) receives input of the results of human judgment made by at least one or more persons and counts the number of people that think it should be retrieved and the number of people that think it should not be retrieved.

[0023]

Relational judgment module (3) determines whether or not the data should be retrieved from the number of people who think it should be retrieved and the number of people who think it should not be retrieved.

[0024]

Constant value generation module (4) generates a predetermined example numerical value when the judgment produced by the retrieval technique and the human judgment agree as determined in external evaluation value and retrieval technique agreement determination module (1).

[0025]

Penalty calculation module (5) determines the penalty value for the fact that the retrieval technique considered a wrong judgment from the ratio of the number of people for whether or not something should be retrieved produced by humans when the judgment produced by the retrieval technique and the human judgment do not agree.

[0026]

Average penalty calculation module (6) finds the average of the penalty values calculated for the combination of all the retrieval conditions and the retrieval data for one retrieval technique.

[0027]

Evaluation value buffer (7) stores the average values calculated as evaluation values along with the retrieval technique.

[0028]

Sorting module (8) rearranges the retrieval techniques in evaluation value buffer (7) in order of the smaller evaluation values.

[0029]

With the application example that is constituted as described above, when the capability of a retrieval technique is measured, a table of the information retrieval technique retrieval results which is the database retrieval results produced with each type of retrieval technique and an external evaluation table that is the result of a judgment as to whether or not something should be retrieved made by a human for the same data are externally input in advance.

[0030]

Of these input data, the table of information retrieval technique retrieval results, as shown in Figure 3, is composed of a technique number for the retrieval technique, an identifier that represents the set of each retrieval condition and each datum in the database, and a judgment flag that represents whether or not said retrieval technique should retrieve the set represented by each identifier.

[0031]

An external evaluation value table, as shown in Figure 4, is composed of an identifier that represents the set of each retrieval condition and each datum in the database, and a judgment flag that indicates whether or not each person thinks each should be retrieved for the set of retrieval conditions and data represented by each identifier.

[0032]

For the process in the first application example, as shown in Figure 5, first, as step 11, the number of people that think it should be retrieved and the number of people that think it should not be retrieved are each counted in external evaluation value totaling module (4) for the data in the external evaluation value table that is the result judged by humans. Next, as step 12, it is judged whether to be data that should be retrieved or data that should not be retrieved from the ratio of the number of people that think it should be retrieved and the number of people that think it should not be retrieved in relational judgment module (3). The value of the judgment result is sent to external evaluation value and retrieval technique agreement determination module (1) along with the number of people that think it should be retrieved and the number of people who think it should not be retrieved, where they are stored in an external evaluation value buffer (shown in Figure 6) provided for external evaluation value and retrieval technique agreement determination module (1).

[0033]

A method for judging whether or not to retrieve at step 12 described above can be realized by comparing the number of people that think something should be retrieved and the number of people who think it should not be retrieved and letting the majority side [decide]. For example, in the case of AAa009 in Figure 4, the number of people who think it should be retrieved is 5, and the number of people who think it should not be retrieved is 2, so it is determined to be data that should be retrieved.

[0034]

As another method for deciding data to be retrieved at step 12, when the judgment from humans is split, such as when the number of people that think AAa002, for example, should be retrieved is 4, and the number of people who think it should not be retrieved is 3, there is the possibility that a large amount of meaning information will be provided. So assuming that such data should not be retrieved, and for example, setting the threshold at 70% beforehand, when following Formula (3) is satisfied,

Formula 2

$$\frac{\text{①} \text{検索べきとした人の数}}{\text{②} \text{検索べきとした人の数} + \text{③} \text{検索べきでないとした人の数}} > 0.7 \quad \dots \text{④}$$

- Key:
- 1 Number of people who think it should be retrieved
 - 2 Number of people who think it should be retrieved
 - 3 Number of people who think it should not be retrieved

[the method] can be realized by determining that it is data that should be retrieved.

[0035]

Next, data in an information retrieval technique retrieval result table given for individual retrieval conditions and individual data in a database are sent to external evaluation value and retrieval technique agreement determination module (1) as step 13 and stored in sequence in an evaluation value storage buffer in the module. The data input from the information retrieval technique retrieval result table and the data sent from relational judgment module (3) are stored together as shown in Figure 6 in the evaluation value storage buffer. When two or more retrieval techniques are evaluated, the input data from the information retrieval technique retrieval result table are replaced and processed in sequence. Next, as step 14, the results of the retrieval technique for each retrieval condition/data identifier and the results from relational judgment module (3) are compared in external evaluation value and retrieval technique agreement determination module (1). When the two values are equal, a signal is sent to constant value generation module (4) as step 15, and a value of 0, for example, is generated and sent to average penalty calculation module (6). When the results of comparison at step 14 is that the retrieval technique and the value of the determination from the relational judgment module are different, the number of people that think it should be retrieved and the number of people who think it should not be retrieved in the evaluation value storage buffer are sent to penalty calculation module (5). As step 16, the penalty when the information retrieval technique and the value of relational judgment module (3) are different is calculated by a predetermined formula in penalty calculation module (5).

[0036]

One actual method of penalty calculation at step 16 can be realized by generating a predetermined constant value. For example, in the case of AAa001 where the value of the information retrieval technique retrieval result and the value from the relational judgment module are different, as shown in Figure 6, 1 is generated as the constant value, for example, by step 16 and this value is sent to average penalty calculation module (6).

[0037]

The average value of the values that indicate the degree of matching between the retrieval technique and human judgment obtained by constant value generation module and penalty calculation module (5) is calculated as step 17 and an evaluation value with the degree of matching of human and retrieval technique for the entire retrieval technique is expressed as a numerical value. Next, as step 18, the evaluation value is stored in evaluation value buffer (7) along with a number assigned to the retrieval technique.

[0038]

The processing between step 13 to step 18 is performed for all given retrieval techniques. When it is determined that processing for all given retrieval techniques is completed at step 18-1, the retrieval technique number and evaluation value are stored in evaluation value buffer (7), and they are rearranged in order of smaller evaluation value in sorting module (8) as step 19. Rearrangement at step 19 can easily be realized with existing technology, such as insertion, heap sort, quick sort, or the like. By outputting the retrieval technique numbers arranged in order of smaller evaluation value as step 20, they can be selected in order from the retrieval techniques closer to human judgment.

[0039]

As another method for realizing penalty calculation at step 16, when AAa100 and AAb001 in Figure 6 are compared, in contrast to the fact that the result from the retrieval technique gives a different judgment, and despite the fact that nearly all the people think that AAa100 should not be retrieved, with AAb001, the number of people that think it should be retrieved and the number of people that think it should not be retrieved are extremely close, so it cannot be said that the judgment by the retrieval technique differs from that of the humans whether to retrieve or not. So by changing the penalty value according to the number of people, giving a larger penalty when there is great variation from the human opinion, and giving a smaller penalty when the human judgment is ambiguous, classification can be accomplished more precisely in order starting with the retrieval techniques closer to human judgment. To realize this, for example, penalty calculation module (5) would be configured as shown in Figure 7.

/8

[0040]

Penalty calculation module (5) shown in Figure 7 is composed of working buffer (21) that temporarily stores the number of people that think something should be retrieved and the number of people that think it should not be retrieved that are input, comparison module (22) that compares

the two data in working buffer (21), division module (23) that divides the compared result using the value of the smaller as the numerator and the value of the larger as the denominator, and penalty value generation module (24) that uses the value of the divided value subtracted from 1 as the penalty value.

[0041]

The processing in penalty calculation module (5) is accomplished as follows. As shown in Figure 8, the number of people that think it should be retrieved and the number of people that think it should not be retrieved are input are stored in working buffer (21) as step 25. Next, as step 28, the values of the two data in working buffer (21) are compared in comparison module (22) and the compared result is stored with the smaller value in numerator buffer (23b) of division module (23) and the larger value in denominator buffer (23a). As step 27, the values in numerator buffer (23b) and denominator buffer (23a) are divided. Finally as step 28, the value divided in penalty value generation module (24) subtracted from 1 is used as the penalty value.

[0042]

With this method of realization, when the judgment by the retrieval technique is different, despite the fact that the human judgments nearly all agree, as for AAa100 in Figure 6, $1 - 1 \div 6 = 0.83$ is given. When the human opinions are split as for AAb001, $1 - 3 \div 4 = 0.25$ is given, which is a smaller penalty value. The result will be that when the human judgment is split, the penalty value for techniques that will give many of the human judgments will be small, and the average penalty value for techniques that show different judgments despite the fact that the human judgments agree will be larger. So both can be clearly separated.

[0043]

The idea of the penalty in this invention is based on the fact that if the result produced by a retrieval technique is to give a different judgment, despite the fact that nearly all the humans think that it should not be retrieved, as for AAa100 in Figure 6, the degree of importance of the information that the retrieval technique does not agree with human judgment is high. If the number of people that think it should be retrieved and the number of people that think it should not be retrieved is very competitive, as for Aab001, it cannot be determined unconditionally whether or not it should be retrieved, so the importance of the information that there is not agreement with human judgment can be considered small.

[0044]

This idea can also be realized with a technique where processing at step 16 applies an entropy formula, since there is a portion similar to the concept of information entropy. One application example where an entropy formula is applied can be realized by the configuration as shown in Figure 9. One example of penalty calculation module (5) shown in Figure 9 is composed of working buffer (31) that temporarily stores the number of people that think it should be retrieved and the number of people that think it should not be retrieved that are input, comparison module (22) that compares the two data in working buffer (31), calculation module (32) that calculates the total number of people that made a judgment from the values of the two data in working buffer (31), entropy calculation module (33) that calculates the entropy value from the number of people that think it should be retrieved, the number of people that think it should not be retrieved, and the number of people responding from obtained from number of people responding calculation module (32), and penalty value generation module (34) that uses the entropy value subtracted from 1 as the penalty value.

[0045]

As processing in the penalty calculation module (5) that is configured as described above, as shown in Figure 10, as step 35, the number of people that think something should be retrieved and the number of people that think it should not be retrieved that are input are stored in working buffer (31). Next, as step 36, the values for the two data in working buffer (31) are added together to calculate the number of people in number of people responding calculation module (32). Next as step 37, entropy calculation is performed with Formula (4) shown below in entropy calculation module (33).

[0046]

Formula 3

$$E = \frac{1}{\log_2 \frac{\text{検索すべきとした人數}}{\text{検索すべきでないとした人數}}} - \frac{1}{\log_2 \frac{\text{検索すべきとした人數}}{\text{検索すべきでないとした人數}}} \dots (4)$$

Key: 1 Number of people that think something should be retrieved
2 Solved number of people

(1-E) is calculated for entropy (E) found here in penalty value generation module (34) as step 38 to accomplish penalty calculation. With this method of realization, when the retrieval technique judgment differs, despite the fact that the judgments of the humans nearly all agree as for AAa100 in Figure 6, a small penalty value of $1 - 0.99 = 0.01$ will be given. The result is that the penalty

value for a technique that differs from many human judgments when the human judgments are split will be smaller. With a technique that shows a different judgment, despite the fact that the human judgments nearly all agree, the average penalty value will be larger, so both can be clearly separated.

[0047]

Next, a second application example of this invention will be explained. Figure 11 is a block diagram of an information retrieval technique evaluation device in a second application example. Said device is composed of a computer, which is primarily a CPU, and programs.

[0048]

With this second application example, as shown in Figure 11, nearly the same configuration as the first application example shown in Figure 1 is used. In this application example, in the information retrieval technique retrieval result table, which is the retrieval result for a database produced by each type of retrieval technique, and an external evaluation value table, which is the result of judgment by humans as to whether or not the same data should be retrieved, are externally input.

[0049]

To measure the capability of a retrieval technique from the input data, the information retrieval technique evaluation device in the second application example is composed of 10 modules: external evaluation value and retrieval technique agreement determination module (1) that compares the judgment as to whether or not something should be retrieved produced by the retrieval technique and the judgments given by humans, external evaluation value totaling module (2) into which the result of human judgment made by at least one or more persons is input and that totals the number of people that think something should be retrieved and the number of people that think it should not be retrieved, (3) relational judgment module (3) that determines whether or not something should be retrieved from the ratio of the number of people that think it should be retrieved and the number of people that think it should not be retrieved, constant value generation module (4) that generates a predetermined constant value when the judgment produced by the retrieval technique and the human judgment agree as determined by external evaluation value and retrieval technique agreement determination module (1), penalty calculation module (5) that determines a penalty value for the fact that the retrieval technique has made an incorrect judgment from the ratio of people that think it should or should not be retrieved when the judgment produced by the retrieval technique and the human judgment do not agree, average penalty calculation module (6) finds the average of the penalty values calculated for combination of all the retrieval

/9

conditions and data to be retrieved for one retrieval technique, actual evaluation value calculation module (41) that calculates the amount of information for human judgments from the ratio of people that think something should or should not be retrieved, imaginary evaluation value calculation module (42) that calculates the amount of information for human judgments for when human judgment changes slightly, intermediate buffer (43) that stores the calculated average penalty value and the actual evaluation value and imaginary evaluation value as a set with the retrieval technique, and technique order calculation module (44) that sequences the retrieval technique in intermediate buffer (43) from the evaluation value, actual evaluation value, and imaginary evaluation value. Here too, the same as in the first application example, each module is also constituted as a program (software) processing unit, but they can also be constituted with hardware.

[0050]

The information retrieval technique retrieval result table and the external evaluation value table, which are input data, are each the same as in the first application example described above, so they will be omitted. However, in the first application example, there was only one external evaluation value table for all the retrieval techniques being evaluated, but in this second application example, an external evaluation value table with a different number of people performing evaluation can also be used for each evaluation technique.

[0051]

As the processing in the second application example, as shown in Figure 2, first, as step 51, for the data in the external evaluation value table, which is the result of judgment by humans, the number of people that think it should be retrieved and the number of people that think it should not be retrieved are each counted in eternal evaluation value totaling module, and [this] is sent to relational judgment module (3) and to actual evaluation value calculation module (41) and imaginary evaluation value calculation module (42).

[0052]

Next as step 52, it is determined in relational judgment module (3) whether it is data that should be retrieved or data that should not be retrieved, from the ratio of the number of people that think it should be retrieved and the number of people that it should not be retrieved. The value for the result of the judgment is sent to external evaluation value and retrieval technique agreement determination module (1) along with the total of people that think it should be retrieved and the people that think it should not be retrieved, where it is stored in the external evaluation value buffer shown in Figure 6.

[0053]

The method for judging whether or not something should be retrieved at step 52 is the same method as used in the first application example, so an explanation will be omitted.

[0054]

Next, data in the conditional retrieval technique retrieval result table given for individual retrieval condition and individual data in the database are sent to external evaluation value and retrieval technique agreement determination module (1) as step 53, where they are stored in order in an evaluation value storage buffer in the module. The evaluation value storage buffer stores data input from the information retrieval technique retrieval result table and data sent from relational judgment module (3) in pairs as shown in Figure 4. When two or more retrieval techniques are evaluated, data input from the retrieval technique retrieval result and the external evaluation value table are replaced and processed in sequence. Next, as step 54, the results of the retrieval technique for each retrieval condition and data identification and the results from relational module (3) are compared in the external evaluation value and retrieval technique determination module. When the two values are equal, as step 55, a signal is sent to constant value generation module (4), a value of 0, for example, is generated, and it is sent to average penalty calculation module (6).

[0055]

When the result of comparison at step 54 is that the value determined by the retrieval technique and by relational judgment module (3) are different, the value for the number of people that think it should be retrieved and the value for the number of people that think it should not be retrieved in the evaluation value storage buffer are sent to penalty calculation module (5). As step 56, a penalty for when the information value from the retrieval technique and from relational judgment module (3) are different is calculated with a predetermined formula in penalty calculation module (5).

[0056]

One method of realizing the penalty calculation at step 56 is to use the technique using entropy in the first application example, for example, so an explanation will be omitted here.

[0057]

For a value that represents the degree of agreement between retrieval technique and human judgments obtained by constant value generation module (4) and penalty calculation module (5), the average value is calculated as step 57, and an evaluation value is generated to express the

degree of agreement between human and retrieval technique judgments for the entire retrieval technique.

[0058]

In actual evaluation value calculation buffer (41), as step 58, the amount of information is calculated from the number of humans that think it should be retrieved and the number of humans that think it should not be retrieved. Calculation of the amount of information can be accomplished with the same technique as calculation of entropy in the first application example, /10 for example, so an explanation will be omitted, since it can easily be analogized.

[0059]

In imaginary evaluation value calculation buffer (42), as step 59, the majority side of the number of people that think something should be retrieved and the number of people that think it should not be retrieved is reduced by a predetermined number of people, for example, 1 person, and the number of people reduced is added to the minority side. So noise can be added to the result of human judgment.

[0060]

Next, as step 60, the amount of information for a case where noise is provided is calculated to give an imaginary evaluation value. Calculation of the imaginary evaluation value is the same as the calculation of entropy in the first application example, so an explanation will be omitted here. Processing in actual evaluation value calculation module (41) and imaginary evaluation value calculation module (42) is performed for all cases as is apparent from Figure 11.

[0061]

Next, as step 61, the evaluation value, actual evaluation value, and imaginary evaluation value are stored in intermediate buffer (43) along with a number assigned to the retrieval technique.

[0062]

The processing between step 51 to step 61 is performed for all the given retrieval techniques. When it is determined at step 61-1 that processing for all given retrieval techniques is completed, the retrieval technique number and evaluation value are stored in intermediate buffer (43). As step 62, the order of the retrieval techniques in the order of similarity to human judgment is determined in technique order calculation module (44) from the evaluation value and the actual

evaluation value and imaginary evaluation value. In addition, as step 63, the technique retrieval numbers are output in order of the smaller evaluation value.

[0063]

Technique order evaluation module (44) is realized with a configuration as shown in Figure 13, for example. Technique order calculation module (44) is constituted from 4 modules: confidence interval calculation module (71) that calculates a confidence interval from the three values of evaluation value, actual evaluation value and imaginary evaluation value sent from intermediate buffer (43), sequencing module (72) that sequences the retrieval techniques relative to the evaluation value and calculated lower and upper limits of each, synthesizing uniform calculation module (73) that generates one order value according to a predetermined formula from the respective order of the lower limit, upper limit and evaluation value for individual retrieval techniques assigned by sequencing module (72), and sorting module (74) that arranges and outputs the order value calculated for each retrieval technique in order from the smallest.

[0064]

As processing in technique order calculation module (44), as shown in Figure 14, as step 81, for example, the lower limit value and upper limit value are calculated using Formulas (5) and (6) below in confidence interval calculation module (71) into which are input the evaluation value and the actual evaluation value and imaginary evaluation value from intermediate buffer (43).

[0065]

Formulas 4

$$\text{下限値} = \text{評価値} - \frac{\text{実測値} - \text{仮想値}}{2} \quad \dots(5)$$

$$\text{上限値} = \text{評価値} + \frac{\text{実測値} - \text{仮想値}}{2} \quad \dots(6)$$

- Key:
- 1 Lower limit value
 - 2 Evaluation value
 - 3 Actual evaluation value
 - 4 Imaginary evaluation value
 - 5 Upper limit value

The calculated upper limit and lower limit value and the evaluation value are stored in a buffer in sequencing module (72) as step 82. Steps 81 and 82 are repeated until calculation of the upper limit

value and lower limit value for all the retrieval techniques is completed. After this, it is determined whether calculation of the upper limit value and lower limit value has been completed for all the retrieval techniques at step 80. After calculation is completed, as step 83, a number is assigned to each technique in order from the smallest for each of the lower limit value, upper limit value, and evaluation value in alignment and order determination part (72a) in sequencing module (72). This processing can easily be realized with a general sorting algorithm, so a detailed explanation will be omitted.

[0066]

Next, as step 84, a synthesized order is calculated from the lower limit order, evaluation order, and upper limit order in synthesized order value calculation module (73). The calculation method for synthesized order can be found by totaling each order as shown in the figure of the buffer in synthesized order calculation module (73) in Figure 13, for example.

[0067]

For the value for each of the lower limit order and upper limit order, the lower limit order represents retrieval efficiency when viewed optimistically, and the upper limit value represents retrieval efficiency when viewed pessimistically. Here, as another calculation method for the synthesized order value, if a pessimistic viewpoint is taken for the retrieval technique so that techniques with a larger upper limit value are discarded, when the total of the lower limit order, the upper limit order, and the evaluation order is found, for example, the value of the upper limit order is added two more times. Then, in the case of Figure 13, technique A will be 13, technique B 8, technique C 9, and technique D 20, and the order value of technique B will be the smallest.

[0068]

Next, as step 85, technique numbers are output in order of the smaller synthesized order values in sorting module (74). Alignment processing of the synthesized order values in sorting module (74) can easily be accomplished with a general sorting algorithm, so a detailed explanation will be omitted.

[0069]

Next, a third application example of this invention will be explained. With the third application example, the same configuration can be added to both the first and the second application examples described above, so here it will be explained using the first application example.

[0070]

A block diagram in the third application example can be represented as shown in Figure 15. In the configuration shown in Figure 15, same number check module (9) and same number occurrence constant generation module (10) are added to the configuration of the first application example.

[0071]

For processing in the third application example, as shown in Figure 16, the number of people that think it should be retrieved and the number of people that think it should not be retrieved are compared at same number check module (9) to determine whether or not they are the same number as step 91 between step 11 and step 21 in Figure 5. Next when the number of people is the same number, as step 92, -1, for example, is sent to external evaluation value and retrieval technique agreement determination module (1), along with the number of people that think it should be retrieved and the number of people that think it should not be retrieved, in place of the results from relational judgment module (9) as impossible to determine.

/11

[0072]

When the number of people in the results for determining whether or not something should be retrieved is not the same, step 12 is executed. Processing in steps 12 and 13 are the same as the first application example, so an explanation will be omitted.

[0073]

Next, in place of step 14 shown in Figure 5, as step 93, the results of the retrieval technique for each retrieval condition and data identifier and the results of relational judgment module (3) are compared in external evaluation value and retrieval technique agreement determination module (1). When the two values are equal, or determination is impossible, for example, when -1 is given to the determination flag produced by relational judgment module (3) in the buffer, as step 15, a signal is sent to same number occurrence constant generation module (10), a value of 0, for example, is generated, and it is sent to average penalty calculation module (6). Processing after step 15 is the same as the first application example, so an explanation will be omitted.

[0074]

When the same function is added to the technique in the second application example, it can be achieved by adding same number check module (9) between external evaluation value totaling module (2) and relational judgment module (3) in Figure 11 in the same way. Processing when it is

added is the same as when same number check module (9) is added to the configuration of the first application example, and an explanation will be omitted, since it can easily be analogized.

[0075]

An example of data in the buffer in external evaluation value and retrieval technique agreement determination module (1) in the third application example is as shown in Figure 17.

[0076]

Next, the results obtained when the capabilities of this invention were tested are shown below.

Experiment conditions:

Number of data to be retrieved used for the experiment: 12 (retrieval objects are composed of a set of keywords accompanying newspaper articles selected at random, and the nouns, proper nouns, and verbal nouns comprising the newspaper article text)

Number of samples of retrieved data actually used: 50 for each condition (randomly extracted)

Retrieval techniques compared: 4

1. AND retrieval: retrieved data acquired with all keywords in the retrieval object

[0077]

2. OR retrieval: retrieved data acquired with any one or more of the keywords in the retrieval object.

[0078]

3. Best retrieval: Retrieval performed artificially so that both the relevance factor and recall ratio will be high for evaluation by combining any one or more keywords in the retrieval object.

4. Single word retrieval: data with a high degree of matching to single words in the set of nouns, proper nouns and verbal nouns comprising the text of the retrieval object is acquired.

[0079]

Number of test subjects: 9

Number of extended one evaluation values: $12 \times 50 = 600$.

Calculation methods in the evaluation: The technique where the concept of entropy was applied in the first through third application examples was used. With this experimental example, the number

of test subjects is an odd number, and the judgments made by human are the same, so the second and third application examples will be the same as the first application example. This is because the width between the upper limit value and the lower limit value in the second application example is the same for any technique, and the human judgments as to whether or not something should be retrieved in the third application example will not be the same number.

[0080]

Results: the results when the experiment was performed are shown below. The relationship between recall ratio and relevance factor used in the past and the evaluation value used in this invention is reflected in the following table.

[0081]

Table 1

	(2)	(3)	(4)
1	検索手法	再現率	適合率
5	AND検索	1.8%	50.0%
6	OR検索	100.0%	36.5%
7	最良検索	48.0%	96.5%
8	單語検索	55.3%	73.0%
			0.145
			0.318
			0.045
			0.076

- Key:
- 1 Retrieval technique
 - 2 Recall ratio
 - 3 Relevance factor
 - 4 This invented technique
 - 5 AND retrieval
 - 6 OR retrieval
 - 7 Best retrieval
 - 8 Single word retrieval

As is clear from the table, using just the recall ratio and relevance factor, it cannot be determined whether OR retrieval is better, or best retrieval is better, or single word retrieval is better. However, in the case of OR retrieval, data that should not be retrieved in the judgment of the most people was included. With this invented technique, there is clearly a difference between OR retrieval, best retrieval, and single word retrieval. In addition, it also shows that the best retrieval that performed a retrieval so that an artificially high retrieval efficiency would occur is the optimum method. The result is that, with this invented technique, it was possible to determine the merits and disadvantages of the retrieval techniques to be, in order starting with the best retrieval technique, best retrieval, single word retrieval, AND retrieval, and OR retrieval.

[0082]

Effect of the invention

As explained above, with the information retrieval technique evaluation method mentioned in Claim 1 of this invention, in a database retrieval technique for acquiring data that matches the retrieval conditions from a database by inputting the retrieval conditions, by inputting judgment data for whether or not retrieval conditions are matched and data for the retrieval results obtained by each retrieval technique for retrieval conditions provided at least one or more humans and for each datum in the database or sample data extracted from the database, the merits and disadvantages of the retrieval efficiency of at least one or more retrieval techniques can be determined automatically.

/12

[0083]

Also, with Claim 2, in addition to the effect described above, a difference occurs for the result of human judgment in the reliability of values viewed statistically when judgments are made by a large number of people and when judgments are made by a small number of people, e.g., 1 or 2. The merits of a retrieval technique can be determined even though the people who are the test subjects in the human based evaluation differ regarding the retrieval technique.

[0084]

Also, with Claim 3, in addition to the effects described above, even if the number of people that make a judgment is an even number and the number of people that think it should be retrieved and the number of people that think it should not be retrieved is the same, by comparing the number of people that think it should be retrieved and the number of people that think it should not be retrieved, between one datum to be retrieved and one [other] datum to be retrieved, and generating a predetermined value when they are the same number, the merits of the retrieval system can be determined when an evaluation is performed with any number of people.

[0085]

Also, with the information retrieval technique evaluation device mentioned in Claim 4, in a database retrieval technique for acquiring data that matches the retrieval conditions from a database by inputting the retrieval conditions, by inputting judgment data for whether or not retrieval conditions are matched and data for the retrieval results obtained by each retrieval technique for retrieval conditions provided by at least one or more humans and for each datum in the database or sample data extracted from the database, the merits and disadvantages of the retrieval efficiency of at least one or more retrieval techniques can be determined automatically.

[0086]

Also, with Claim 5, in addition to the effect described above, a difference occurs for the result of human judgment in the reliability of values viewed statistically when judgments are made by a large number of people and when judgments are made by a small number of people, e.g., 1 or 2. The merits of a retrieval technique can be determined even though the people who are the test subjects in the human based evaluation differ regarding to the retrieval technique.

[0087]

Also, with Claim 6, in addition to the effects described above, even if the number of people that make a judgment is an even number and the number of people that think it should be retrieved and the number of people that think it should not be retrieved is the same, by comparing the number of people that think it should be retrieved and the number of people that think it should not be retrieved, between one datum to be retrieved and one [other] datum to be retrieved, and generating a predetermined value with a constant generation means when they are the same number, the merits of the retrieval system can be determined when an evaluation is performed with any number of people.

Brief description of the figures

Figure 1 is a system block diagram of an information retrieval technique evaluation device in a first application example of this invention.

Figure 2 is a conceptual diagram that shows an evaluation method for retrieval efficiency.

Figure 3 is a figure that shows the data structure and data examples in the information retrieval technique retrieval result table in the first application example.

Figure 4 is a figure that shows the data structure and data examples in the external value evaluation table in the first application example.

Figure 5 is a flowchart that shows processing in the information retrieval technique evaluation method in the first application example.

Figure 6 is a figure that shows the data structure in the external evaluation value buffer in the first application example.

Figure 7 is a system block diagram of the penalty calculation module in the first application example.

Figure 8 is a flowchart that shows the flow of processing by the penalty calculation module in the first application example.

Figure 9 is a system block diagram that shows another example of the penalty calculation module.

Figure 10 is a flowchart that shows the processing flow in the other example of the penalty calculation module.

Figure 11 is a system block diagram of an information retrieval technique evaluation device in a second application example.

Figure 12 is a flowchart that shows the processing flow by the information retrieval technique evaluation method in the second application example.

Figure 13 is a system block diagram of the technique order calculation module in the second application example.

Figure 14 is a flowchart that shows the processing in the technique order calculation module in the second application example.

Figure 15 is a system block diagram of an information retrieval technique evaluation device in the third application example.

Figure 16 is a flowchart that shows the processing in the information retrieval technique evaluation method in the third application example.

Figure 17 is a figure that shows the data structure and data examples in the external evaluation value buffer in the third application example.

Explanation of reference symbols

(1) ... external evaluation value and retrieval technique agreement determination module, (2) ... external evaluation value totaling module, (3) ... relational judgment module, (4) ... constant value generation module, (5) ... penalty calculation module, (6) ... average penalty calculation module, (7) ... evaluation value buffer, (8) ... sorting module, (9) ... same number check module, (21) ... working buffer, (22) ... comparison module, (23) ... division module, (24) ... penalty value generation module, (31) ... working buffer, (32) ... number of people responding calculation module, (33) ... entropy calculation module, (34) ... penalty value generation module, (41) ... actual evaluation value calculation module, (42) imaginary evaluation value calculation module, (43) ... intermediate buffer, (44) ... technique order calculation module, (71) ... confidence interval calculation module, (72) ... sequencing module, (73) ... synthesized order value calculation module, (74) ... sorting module.

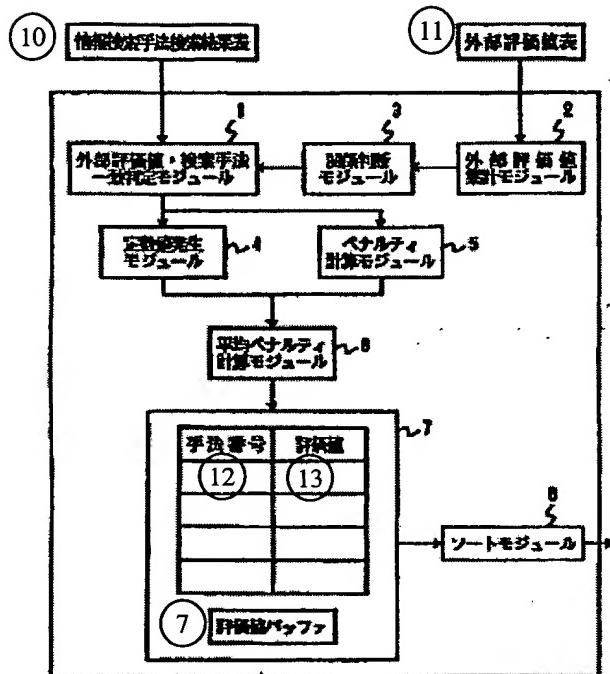


Figure 1

- Key:
- 1 External evaluation value and retrieval technique agreement determination module
 - 2 External evaluation value totaling module
 - 3 Relational judgment module
 - 4 Constant value generation module
 - 5 Penalty calculation module
 - 6 Average penalty calculation module
 - 7 Evaluation value buffer
 - 8 Sorting module
 - 10 Information retrieval technique search result table
 - 11 External evaluation value table
 - 12 Technique number
 - 13 Evaluation value

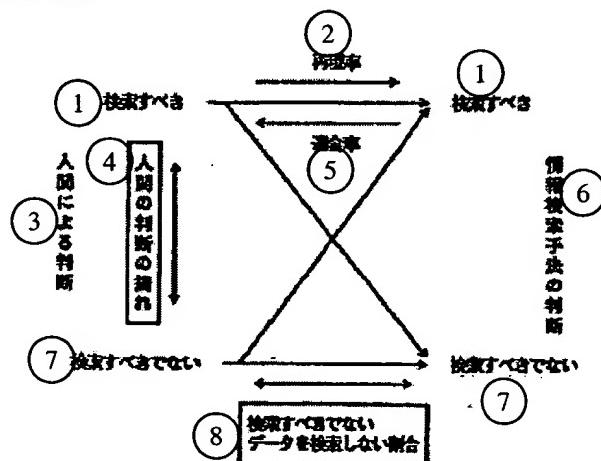


Figure 2

- Key:
- 1 Should be retrieved
 - 2 Recall ratio
 - 3 Judgment by human
 - 4 Wobble in human judgment
 - 5 Relevance factor
 - 6 Judgment by information retrieval technique
 - 7 Should not be retrieved
 - 8 Rate at which data that should not be retrieved are not retrieved

1	AND検索
AAa001	0
AAa002	0
AAa003	1
AAa004	0
:	:
AAa009	0
AAa100	1
AAa101	0
AAb001	1
AAb002	0
AAb003	0
:	:

Figure 3

- Key:
- 1 Retrieval technique name
 - 2 Retrieval technique number
 - 3 Retrieval conditions -- data identifier
 - 4 Judgment flag
(0: should not be retrieved)
(1: should be retrieved)
 - 5 AND retrieval
 - 6 Information retrieval technique retrieval result table

	1	2	3	4	5	6	7
検索条件 1の条件 2の条件 3の条件 4の条件 5の条件 6の条件 7の条件	S	S	S	S	S	S	S
AAa001	1	1	0	1	0	1	1
AAa002	0	0	1	0	0	0	0
AAa003	1	1	1	1	0	1	1
AAa004	0	0	0	0	0	0	0
:	:	:	:	:	:	:	:
AAa009	1	1	1	0	0	1	1
AAa100	0	0	0	1	0	0	0
AAa101	0	0	0	0	1	0	0
ABb001	0	1	1	0	0	1	0
ABb002	0	0	1	1	1	0	1
ABb003	0	0	0	0	0	0	0
:	:	:	:	:	:	:	:

8 検索条件・データ識別子 9 判断フラグ
(0:検索すべきでない、1:検索すべき)
10 外部評価値表

Figure 4

- Key:
- 1 Response of test subject 1
 - 2 Response of test subject 2
 - 3 Response of test subject 3
 - 4 Response of test subject 4
 - 5 Response of test subject 5
 - 6 Response of test subject 6
 - 7 Response of test subject 7
 - 8 Retrieval conditions -- data identifier
 - 9 Judgment flag
(0: should not be retrieved 1: should be retrieved)
 - 10 External evaluation value table

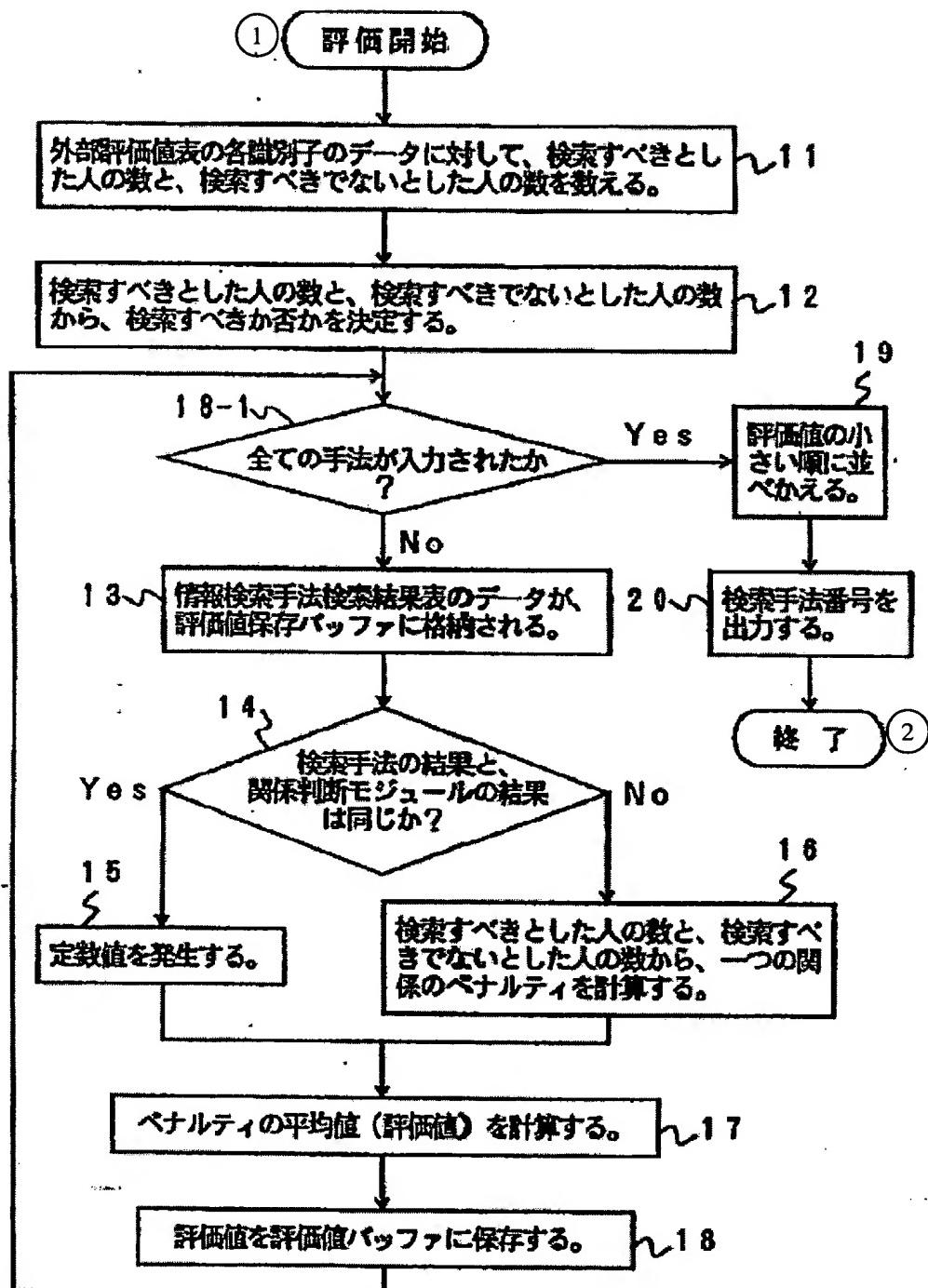


Figure 5

Key: 1 Evaluation start

2 End

11 Count the number of people that think it should be retrieved and the number of people that think it should not be retrieved, for the data for each identifier in external evaluation value table

- 12 Determine whether it should be retrieved from the number of people that think it should be retrieved and the number of people that think it should not be retrieved
- 13 Data in information retrieval technique retrieval result table stored in evaluation value storage buffer
- 14 Retrieval technique result and relational judgment module result the same?
- 15 Constant value is generated
- 16 One relational penalty is calculated from the number of people that think it should be retrieved and the number of people that think it should not be retrieved
- 17 Average value (evaluation value) of penalties calculated
- 18 Evaluation value stored in evaluation value buffer
- 18-1 All techniques input?
- 19 Evaluation values rearranged in order of smaller
- 20 Retrieval technique number is output.

AAa001	0	1	5	2
AAa002	0	0	1	8
AAa003	1	1	0	1
AAa004	0	0	0	7
:	:	:	:	:
AAa099	0	1	5	2
AAa100	1	0	1	6
AAa101	0	0	5	1
AAb001	1	0	3	4
AAb002	0	1	4	3
AAb003	0	0	0	7
:	:	:	:	:

説明欄
レーティング子による判断フラグ
リテラルによる判断フラグ
数値がべきで
した数 4
ないとした人数
5
6

Figure 6

- Key:
- 1 Retrieval conditions – identifier
 - 2 Judgment flag produced by information retrieval technique retrieval result table
 - 3 Judgment flag produced by relational judgment module
 - 4 Number of people that think it should be retrieved
 - 5 Number of people that think it should not be retrieved
 - 6 External evaluation value buffer

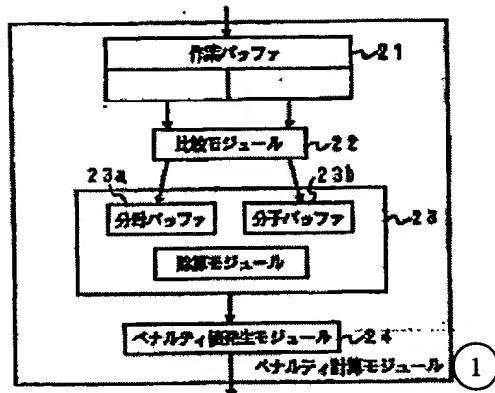


Figure 7

- Key:
- 1 Penalty calculation module
 - 21 Working buffer
 - 22 Comparison module
 - 23 Division module
 - 23a Numerator buffer
 - 23b Denominator buffer
 - 24 Penalty value generation module

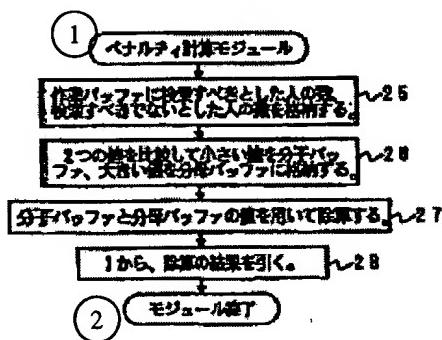


Figure 8

- Key:
- 1 Penalty calculation module
 - 2 Module end
 - 25 Number of people that think it should be retrieved and number of people that think it should not be retrieved stored in working buffer
 - 26 Two values compared, smaller value stored in numerator buffer and larger value stored in denominator buffer
 - 27 Division performed using values in numerator buffer and denominator buffer.
 - 28 Result of division subtracted from 1.

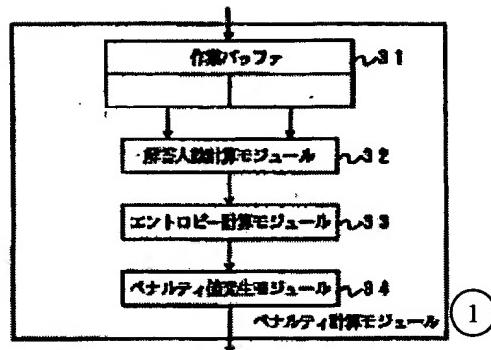


Figure 9

- Key:
- 1 Penalty calculation module
 - 31 Working buffer
 - 32 Number of people responding calculation module
 - 33 Entropy calculation module
 - 34 Penalty value generation module

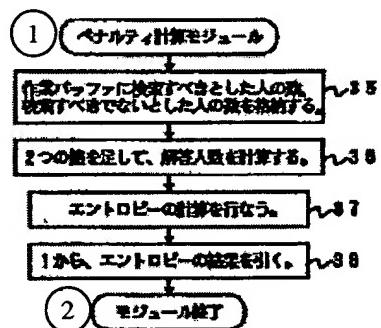


Figure 10

- Key:
- 1 Penalty value calculation module
 - 2 Module end
 - 35 Number of people that think it should be retrieved and number of people that think it should not be retrieved are stored in working buffer
 - 36 Two values are added to calculate number of people responding.
 - 37 Entropy calculated
 - 38 Result of entropy subtracted from 1

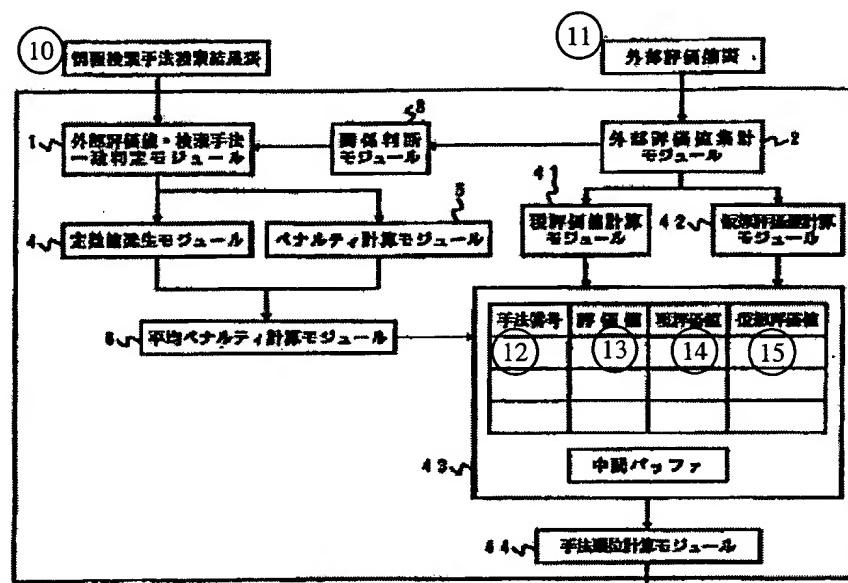


Figure 11

- Key:
- 1 External evaluation value and retrieval technique agreement determination module
 - 2 External evaluation value totaling module
 - 3 Relational judgment module
 - 4 Constant value generation module
 - 5 Penalty calculation module
 - 6 Average penalty calculation module
 - 10 Information retrieval technique retrieval result table
 - 11 External evaluation value table
 - 12 Technique number
 - 13 Evaluation value
 - 14 Actual evaluation value
 - 15 Imaginary evaluation value
 - 41 Actual evaluation value calculation module
 - 42 Imaginary evaluation value calculation module
 - 43 Intermediate buffer
 - 44 Technique order calculation module

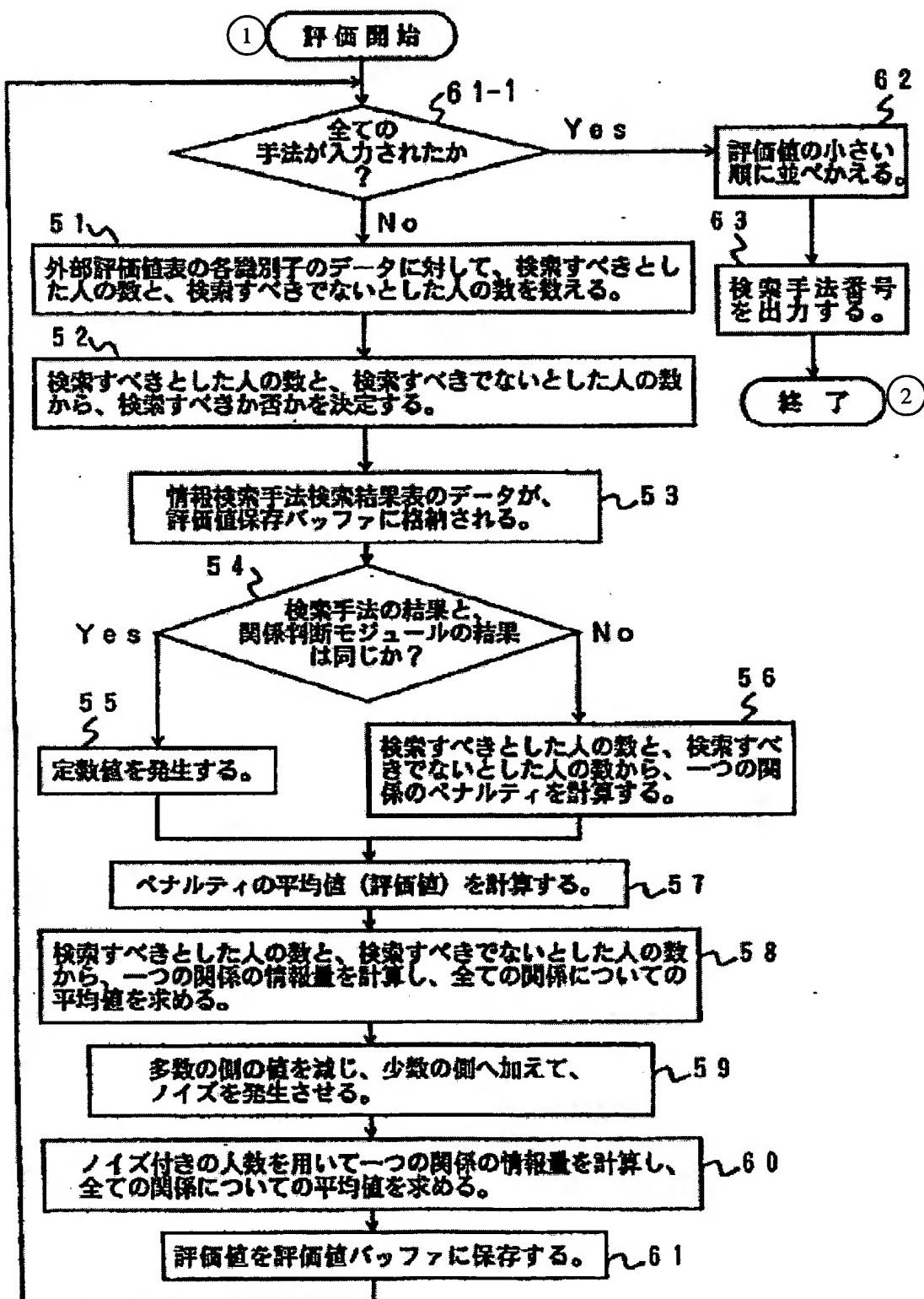


Figure 12

Key:

- 1 Evaluation start
- 2 End

- 51 Number of people that think it should be retrieved and number of people that think it should not be retrieved counted for data for each identifier in external evaluation value table
- 52 Determine whether it should be retrieved from the number of people that think it should be retrieved and number of people that think it should not be retrieved
- 53 Data in information retrieval technique retrieval result table stored in evaluation value storage buffer.
- 54 Result of retrieval technique and result of relational judgment module the same?
- 55 Constant value generated
- 56 One relational penalty calculated from the number of people that think it should be retrieved and number of people that think it should not be retrieved
- 57 Average value (evaluation value) of penalties calculated
- 58 Amount of information for one relation calculated from the number of people that think it should be retrieved and number of people that think it should not be retrieved, and the average value for all relations found
- 59 Value on the majority side is decreased and added to the minority side to generate noise.
- 60 Amount of information for one relation is calculated using the number of people with noise, and the average value for all relations is found.
- 61 Evaluation value stored in evaluation value buffer
- 61-1 All techniques input?
- 62 Rearranged in order of smaller evaluation value

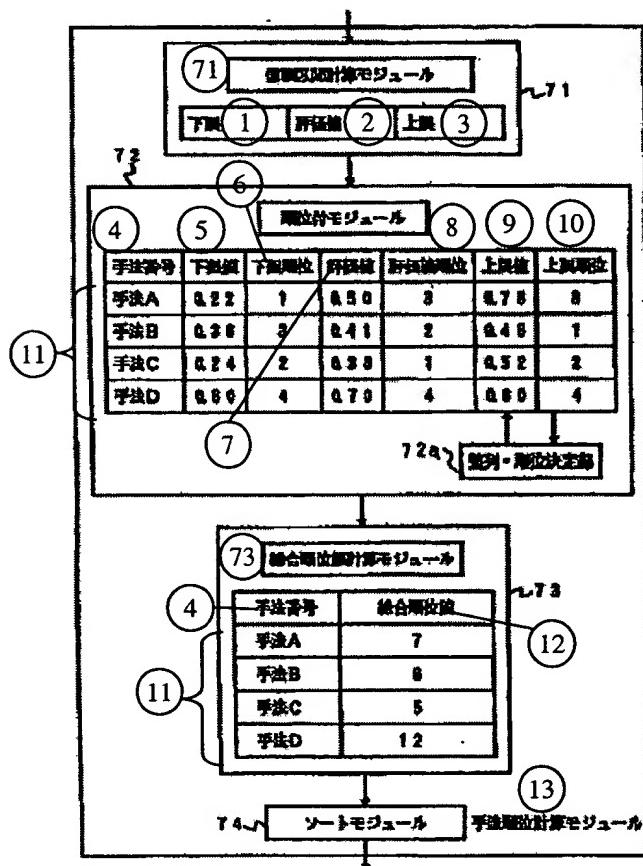


Figure 13

- Key:
- 1 Lower limit
 - 2 Evaluation value
 - 3 Upper limit
 - 4 Technique number
 - 5 Lower limit value
 - 6 Lower limit order
 - 7 Evaluation value
 - 8 Evaluation value order
 - 9 Upper limit value
 - 10 Upper limit order
 - 11 Technique [A, B, etc.]
 - 12 Synthesized order value
 - 13 Technique order calculation module
 - 71 Confidence interval calculation module
 - 72a Arrangement and order determination part
 - 73 Synthesized order value calculation module
 - 74 Sorting module

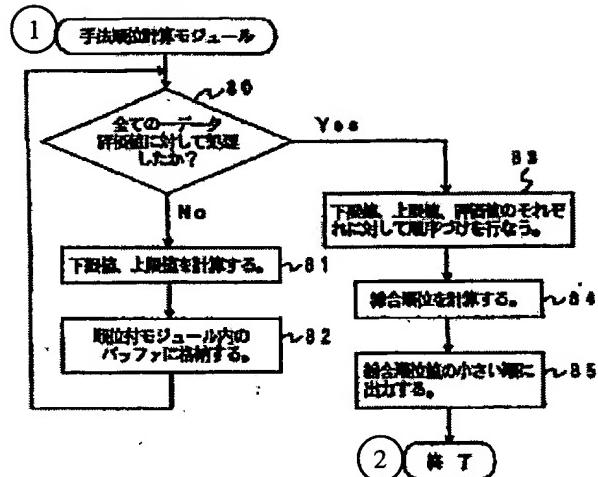


Figure 14

- Key:
- 1 Technique order calculation module
 - 2 End
 - 80 Processing done for all data evaluation values?
 - 81 Lower limit value and upper limit value calculated
 - 82 Stored in buffer in sequencing module
 - 83 Sequencing performed for lower limit value, upper limit value, and evaluation value
 - 84 Synthesized order calculated
 - 85 Output in order of smaller synthesized order value.

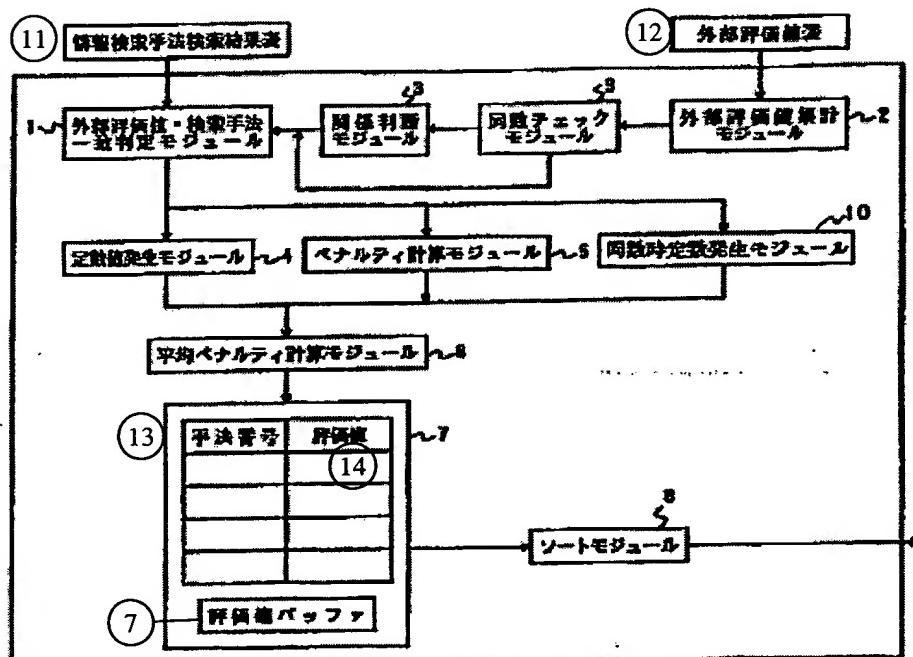


Figure 15

- Key:
- 1 External evaluation value and retrieval technique agreement determination module
 - 2 External evaluation value totaling module
 - 3 Relational judgment module
 - 4 Constant value generation module
 - 5 Penalty calculation module
 - 6 Average penalty calculation module
 - 7 Evaluation value buffer
 - 8 Sorting module
 - 9 Same number check module
 - 10 Same number occurrence constant generation module
 - 11 Information retrieval technique retrieval result table
 - 12 External evaluation value table
 - 13 Technique number
 - 14 Evaluation value

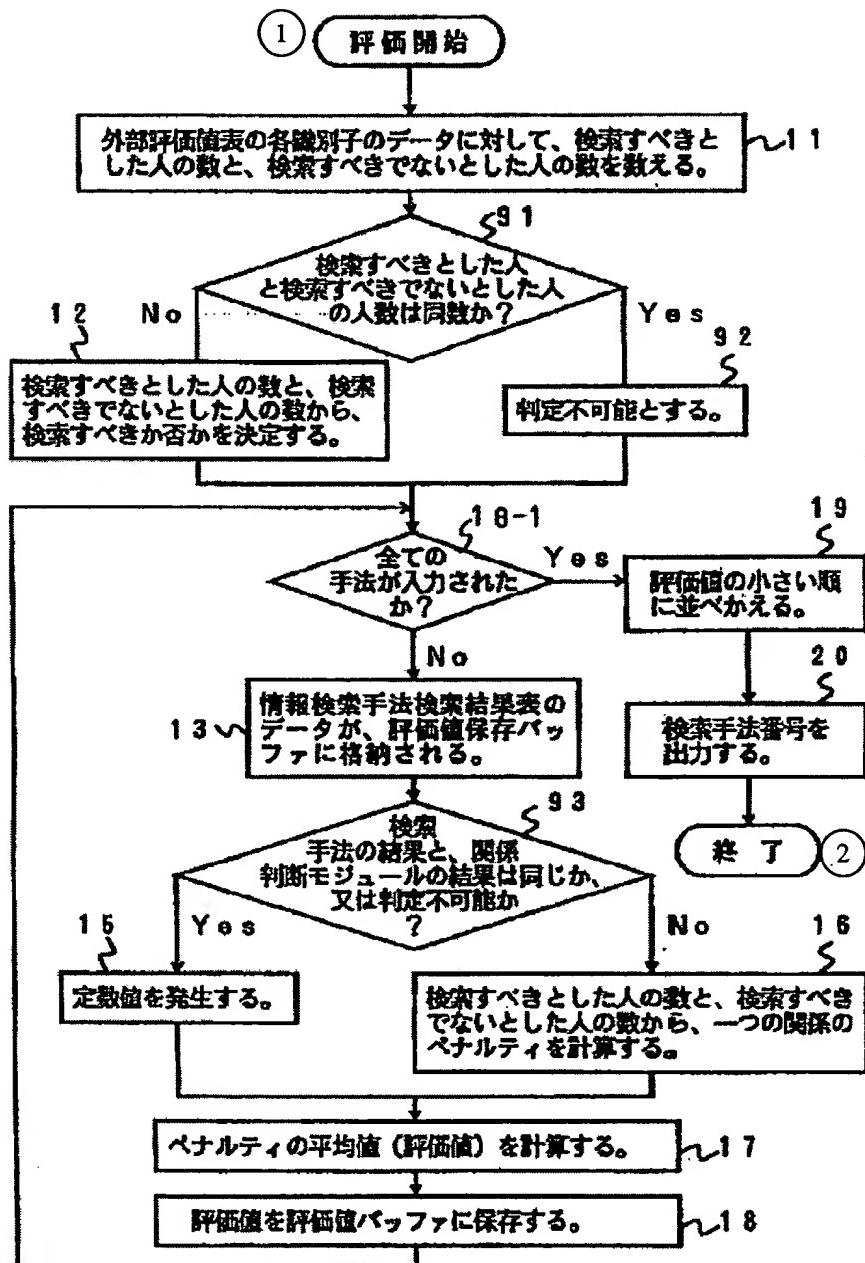


Figure 16

- Key:
- 1 Evaluation start
 - 2 End
 - 11 Number of people that think it should be retrieved and number of people that think it should not be retrieved counted for data for each identifier in external evaluation value table
 - 12 Determine whether it should be retrieved from the number of people that think it should be retrieved and number of people that think it should not be retrieved
 - 13 Data in information retrieval technique retrieval result table stored in evaluation value storage buffer

- 15 Constant value is generated
 - 16 One relational penalty is calculated from the number of people that think it should be retrieved and the number of people that think it should not be retrieved
 - 17 Average value (evaluation value) of penalties calculated
 - 18 Evaluation value stored in evaluation value buffer
 - 18-1 All techniques input?
 - 19 Rearranged in order of smaller evaluation value
 - 20 Retrieval technique number is output
 - 91 Number of people that think it should be retrieved and number of people that think it should not be retrieved the same number?
 - 92 Decision that it is not possible to determine
 - 93 Result of retrieval technique and result of relational judgment module the same, or determination impossible?

AAa001	0	1	5	2
AAa002	0	0	1	5
AAa003	1	1	6	1
AAa004	0	0	0	7
:	:	:	:	:
AAa005	0	1	5	2
AAa100	1	0	1	5
AAa101	0	0	0	1
AAb001	1	0	3	4
AAb002	0	1	4	3
AAb003	0	0	0	7
:	:	:	:	:

検査条件 - 情報検索文字検索結果
データ属性ID がによる判断フラグ

検査すべきとした人選 4 確認すべきでないとした人無

① ② ③ ④ ⑤ ⑥

情報検索モジュールによる判断フラグ
登録詳細バッファ

Figure 17

- | | | |
|------|---|--|
| Key: | 1 | Retrieval conditions – identifier |
| | 2 | Judgment flag produced by information retrieval technique retrieval result table |
| | 3 | Judgment flag produced by relational judgment module |
| | 4 | Number of people that think it should be retrieved |
| | 5 | Number of people that think it should not be retrieved |
| | 6 | External evaluation value buffer |

(19) 日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平7-271798

(43) 公開日 平成7年(1995)10月20日

(51) Int.Cl.^a
G 0 6 F 17/30識別記号
9194-5L

F I

G 0 6 F 15/ 403

3 4 0 B

技術表示箇所

審査請求 未請求 請求項の数 6 OL (全 19 頁)

(21) 出願番号

特願平6-60682

(22) 出願日

平成6年(1994)3月30日

(71) 出願人

000004226
日本電信電話株式会社

東京都千代田区内幸町一丁目1番6号

(72) 発明者 小澤 英昭

東京都千代田区内幸町1丁目1番6号 日
本電信電話株式会社内

(72) 発明者 中川 透

東京都千代田区内幸町1丁目1番6号 日
本電信電話株式会社内

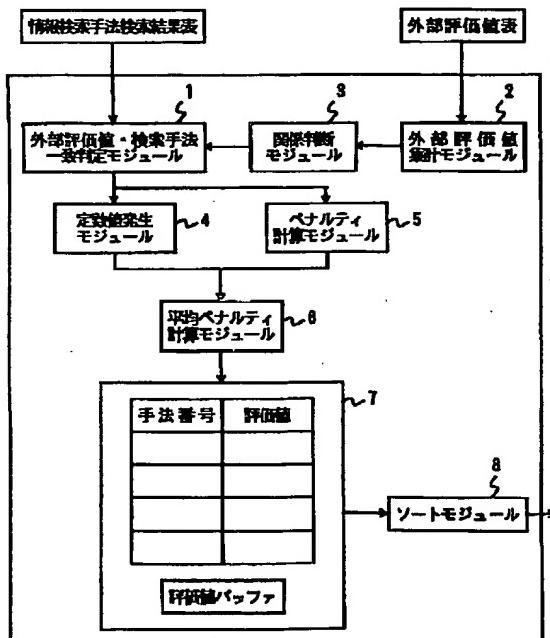
(74) 代理人 弁理士 吉田 精孝

(54) 【発明の名称】 情報検索手法評価方法及びその装置

(57) 【要約】

【目的】 最適な情報検索手法を確実に判断できる情報検索手法評価方法及びその装置を提供する。

【構成】 外部評価値の個数を計測し、この一つのデータに対する情報検索手法の判断が、与えられた外部評価値による判断の多数の側と同じであるならば、定数値発生モジュール4により発生した定数値を一データ評価値とし、小数の側と同じであるならば、ペナルティ計算モジュール5により得られた値を一データ評価値とし、与えられた全ての検索対象データとデータベース中のデータの組合せについて前記一データ評価値を生成して、平均ペナルティ計算モジュール6により、全ての一データ評価値の平均値を計算し、これを情報検索手法の検索効率の評価値として検索効率を一つの指標を用いて表現し、各情報検索手法に対して計算した評価値を順序づけることにより、各種情報検索手法の優劣を判定する。



1

【特許請求の範囲】

【請求項1】 予め与えられた検索対象データに合致する被検索データをデータベース中から獲得する情報検索システム上で利用される情報検索手法の検索効率の評価方法であり、少なくとも一人以上の人間によって判定されたデータベース中のデータが検索対象のデータに合致するか否かを示す外部評価値と、情報検索手法によって判定されたデータベース中のデータが検索対象のデータに合致するか否かの値とから、それぞれの検索手法の検索効率を計算し、検索手法の優劣を判定する情報検索手法評価方法において、

検索対象データとデータベース中の任意の一つのデータが、少なくとも一人以上の人間によって合致すると判定された外部評価値の個数と、合致しないと判断された外部評価値の個数とをそれぞれ計測し、この一つのデータに対する情報検索手法が判定した検索対象データに合致するか否かの判断が、与えられた外部評価値による検索条件に合致するか否かの判断の多数の側と同じであるならば、予め与えられた定数値を一データ評価値とし、小数の側と同じであるならば、多数の側と小数の側の外部評価値の個数から、予め与えられた式によって得られた値を一データ評価値とし、

与えられた全ての検索対象データとデータベース中のデータの組合せについて前記一データ評価値を生成し、全ての一データ評価値の平均値を計算し、該平均値を情報検索手法の検索効率の評価値として検索効率を一つの指標を用いて表現し、少なくとも一つ以上の情報検索手法に対して計算した評価値を順序づけることにより、各種情報検索手法の優劣を判定することを特徴とする情報検索手法評価方法。

【請求項2】 前記一データ評価値を計算する際に、予め定められた式に従って、外部評価値の小数の側の人数と多数の側の人数の値から現評価値を計算し、多数の側の人数から予め定められた人数を減じ、小数の側の人数に多数側から減じた値を加えて、予め定められた式に従い仮想評価値を計算し、

仮想評価値と現評価値との差を一データ評価値の揺れ度合いとし、全ての検索条件とデータベース中のデータの組み合わせに対する揺れ度合いの平均値を計算し、

前記検索効率の値に揺れ度合いの値を増減した信頼区間付き検索効率を計算し、

該信頼区間付き検索効率の最大値と最小値及び前記一データ評価値から予め与えられた式により各検索手法に対して一つの検索効率の値を計算することにより、複数の情報検索手法の検索効率の優劣を判断することを特徴とする請求項1記載の情報検索手法評価方法。

【請求項3】 前記一データ評価値の計算において、外部評価値で検索対象データに合致するに該当する人数と、合致しないに該当する人数とを比較し、これらが同数であった場合には、予め定められた値を一データ評価

2

値として、各検索手法の検索効率を計算することを特徴とする請求項1または2記載の情報検索手法評価方法。

【請求項4】 予め与えられた検索対象データに合致する被検索データをデータベース中から獲得する情報検索システム上で利用される情報検索手法の検索効率の評価装置であり、少なくとも一人以上の人間によって判定されたデータベース中のデータが検索対象のデータに合致するか否かを示す外部評価値と、情報検索手法によって判定されたデータベース中のデータが検索対象のデータに合致するか否かの値とから、それぞれの検索手法の検索効率を計算し、検索手法の優劣を判定する情報検索手法評価装置であって、

少なくとも一人以上の人間より、検索すべきであるのか否かの判断の結果を入力する入力手段と、

該入力された人間の判断結果の中から、一つ一つの検索条件と一つ一つの検索対象データとの間で、検索すべきだとした人数、検索すべきではないとした人数をそれぞれ計測する計測手段と、

該計測手段によって計測された検索すべきだとした人数と、検索すべきではないとした人数とを比較して、予め定められた割合以上に検索すべきだとした人数が多い場合に、該検索対象データは検索すべきだと決定する決定手段と、

前記検索条件と検索対象データに対して、情報検索手法が行った検索すべきか否かの判断が、一致しているか、一致していないかを判断する判断手段と、

該判断の結果、一致している場合には予め定められた定数値を発生する定数値発生手段と、

前記判断手段による判断の結果、一致しない場合には予め定められた式に従って、ペナルティの値を発生するペナルティ値発生手段と、

与えられた全ての検索条件と検索対象データの組み合わせに対して得られた定数値及びペナルティの値の平均値を評価値として求める平均値算出手段と、

各検索手法に対して得られた評価値を所定の順序に並べるソート手段とを設けたことを特徴とする情報検索手法評価装置。

【請求項5】 人間の判断の結果を検索すべきとした人数と、検索すべきでないとした人数を計測した後に、予め定められた式により人間の判断の違いの大きさを計算する計算手段と、

人間の判断にノイズを与えるために、検索すべきとした人数と、検索すべきでないとした人数を比較する比較手段と、

該比較手段により人数を比較した結果、多数の側の人数の値から予め定められた定数値を減じ、小数の側の人数の値に予め定められた定数値を加える加減算手段と、

該加減算手段によって再設定された検索すべきとした人数、及び検索すべきでないとした人数から、予め定められた式に従って再度人間の判断の違いの大きさを計算す

10

20

30

40

50

る再計算手段と、

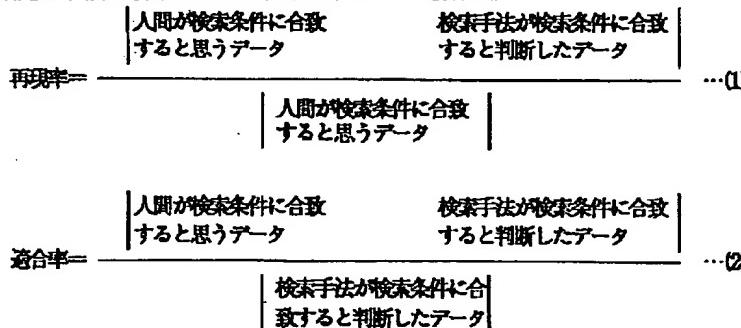
全ての検索条件と検索データの組み合わせに対して、前記再計算手段による人間の判断とノイズの載った判断との違いの大きさの計算を行った結果の平均値を評価値として求める第2の平均値算出手段と、前記平均値算出手段によって得られた評価値と前記第2の平均値算出手段によって得られた評価値とから予め定められた式により、評価値の最小値、最大値を計算する最小最大値算出手段と、

評価値、最小値、最大値について、小さい順に並べることで各検索手法のそれぞれの順位を計算する順位算出手段と、

評価値、最小値、最大値の3つの値それぞれについての各検索手法の順位の値を予め定められた式により一つの値とする順位決定手段とを設けたことを特徴とする請求項4記載の情報検索手法評価装置。

【請求項6】一つの検索対象データと一つの検索対象データとの間で、検索すべきとした人数と、検索すべきでないとした人数を比較する比較手段と、

該比較手段の比較結果が同数であったならば予め定められた値を発生する定数発生手段を設けたことを特徴とす*



【0003】

【発明が解決しようとする課題】しかし、再現率と適合率による指標は分子の値は同じであり、分母の値のみの違いが指標に影響する。分母の値は、再現率では人間の判断による検索すべきデータの個数であり、適合率の場合では検索手法の判断による検索すべきとしたデータの個数である。人間の判断を変えないで再現率を上げようとして検索手法の条件を弛めると、多くの結果が検索されてしまうため適合率が低下し、逆に適合率を上げようとすると、検索手法の条件を厳しくすることになるため、再現率、適合率には、トレードオフの関係がある。このため各種の検索手法の善し悪しを評価する際に、適合率の高い手法をとるべきか、再現率の高い手法をとるべきかの判断により、どの手法が適しているのかが判断できなかった。

【0004】特に適合率は、分母に「検索手法による値」をとっているために、例えば同じ30%という値をとったとしても、人間が判断して9個の正解があるとき※50

* る請求項4又は5記載の情報検索手法評価装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、複数の情報検索手法が存在したときに、最適な情報検索手法を自動的に選択するための、情報検索手法評価方法及びその装置に関するものであり、特に、新聞記事等の時事情報のデータベースに代表される、検索条件と検索すべきデータとの関係が、ユーザーにより変化する場合が多いデータベースの検索手法の評価方法及びその装置に関するものである。

【0002】

【従来の技術】従来、情報検索手法の評価においては、次の(1)式で表される再現率と(2)式で表される適合率と呼ばれる、2つの指標を組み合わせて判断してきた。

(例えばGerard Salton (Ed.), The SMART Retrieval System - Experiments in Automatic Document Processing, Englewood Cliffs, NJ, Prentice-Hall, 1971, B.Masand, G.Linoff, D.Waltz, Classifying News Stories using Memory Based Reasoning. Proceedings, 15th International SIGIR 1992, pp.59-56, 1992等多数)

【数1】

検索手法が検索条件に合致すると判断したデータ

... (1)

検索手法が検索条件に合致すると判断したデータ

... (2)

※に、検索した結果として30個のデータを得て、人間が検索すべきと考える9個全てを含む場合と、10個のデータしか検索せずに3個の正解データを含む場合で、どちらの場合が良い検索手法なのかは、判断できない。

【0005】また再現率においても、時事情報のように検索する人によって、同じ検索対象データから検索を行ったとしても、得たい検索結果の内容が異なる場合には、再現率の分母が変化てしまい、再現率の値が変化してしまう問題点があった。

【0006】この問題点は、検索効率の評価法を図2のように表現した場合に、人間が検索すべきとしたデータが情報検索手法によって検索すべきとなる場合しか注目しておらず、人間の判断の揺れと、検索すべきでないデータを検索しないということが、評価手法に入ってしまひなかつたことに起因する。

【0007】本発明の目的は上記の問題点に鑑み、最適な情報検索手法を確実に判断できる情報検索手法評価方法及びその装置を提供することにある。

【0008】

【課題を解決するための手段】本発明は上記の目的を達成するために請求項1では、予め与えられた検索対象データに合致する被検索データをデータベース中から獲得する情報検索システム上で利用される情報検索手法の検索効率の評価方法であり、少なくとも一人以上の人間によって判定されたデータベース中のデータが検索対象のデータに合致するか否かを示す外部評価値と、情報検索手法によって判定されたデータベース中のデータが検索対象のデータに合致するか否かの値とから、それぞれの検索手法の検索効率を計算し、検索手法の優劣を判定する情報検索手法評価方法において、検索対象データとデータベース中の任意の一つのデータが、少なくとも一人以上の人間によって合致すると判定された外部評価値の個数と、合致しないと判定された外部評価値の個数とをそれぞれ計測し、この一つのデータに対する情報検索手法が判定した検索対象データに合致するか否かの判断が、与えられた外部評価値による検索条件に合致するか否かの判断の多数の側と同じであるならば、予め与えられた定数值を一データ評価値とし、小数の側と同じであるならば、多数の側と小数の側の外部評価値の個数から、予め与えられた式によって得られた値を一データ評価値とし、与えられた全ての検索対象データとデータベース中のデータの組合せについて前記一データ評価値を生成し、全ての一データ評価値の平均値を計算し、該平均値を情報検索手法の検索効率の評価値として検索効率を一つの指標を用いて表現し、少なくとも一つ以上の情報検索手法に対して計算した評価値を順序づけることにより、各種情報検索手法の優劣を判定する情報検索手法評価方法を提案する。

【0009】また、請求項2では、請求項1記載の情報検索手法評価方法において、前記一データ評価値を計算する際に、予め定められた式に従って、外部評価値の小数の側の人数と多数の側の人数の値から現評価値を計算し、多数の側の人数から予め定められた人数を減じ、小数の側の人数に多数側から減じた値を加えて、予め定められた式に従い仮想評価値を計算し、仮想評価値と現評価値との差を一データ評価値の揺れ度合いとし、全ての検索条件とデータベース中のデータの組み合わせに対する揺れ度合いの平均値を計算し、前記検索効率の値に揺れ度合いの値を増加減した信頼区間付き検索効率を計算し、該信頼区間付き検索効率の最大値と最小値及び前記一データ評価値から予め与えられた式により各検索手法に対して一つの検索効率の値を計算することにより、複数の情報検索手法の検索効率の優劣を判断する情報検索手法評価方法を提案する。

【0010】また、請求項3では、請求項1又は2記載の情報検索手法評価方法において、前記一データ評価値の計算において、外部評価値で検索対象データに合致するに該当する人数と、合致しないに該当する人数と比

10

20

30

40

50

較し、これらが同数であった場合には、予め定められた値を一データ評価値として、各検索手法の検索効率を計算する情報検索手法評価方法を提案する。

【0011】また、請求項4では、予め与えられた検索対象データに合致する被検索データをデータベース中から獲得する情報検索システム上で利用される情報検索手法の検索効率の評価装置であり、少なくとも一人以上の人間によって判定されたデータベース中のデータが検索対象のデータに合致するか否かを示す外部評価値と、情報検索手法によって判定されたデータベース中のデータが検索対象のデータに合致するか否かの値とから、それぞれの検索手法の検索効率を計算し、検索手法の優劣を判定する情報検索手法評価装置であって、少なくとも一人以上の人間より、検索すべきであるのか否かの判断の結果を入力する入力手段と、該入力された人間の判断結果の中から、一つ一つの検索条件と一つ一つの検索対象データとの間で、検索すべきだとした人数、検索すべきではないとした人数をそれぞれ計測する計測手段と、該計測手段によって計測された検索すべきだとした人数と、検索すべきではないとした人数とを比較して、予め定められた割合以上に検索すべきだとした人数が多い場合に、該検索対象データは検索すべきだと決定する決定手段と、前記検索条件と検索対象データに対して、情報検索手法が行った検索すべきか否かの判断が、一致しているか、一致していないかを判断する判断手段と、該判断の結果、一致している場合には予め定められた定数值を発生する定数值発生手段と、前記判断手段による判断の結果、一致しない場合には予め定められた式に従って、ペナルティの値を発生するペナルティ値発生手段と、与えられた全ての検索条件と検索対象データの組み合わせに対して得られた定数值及びペナルティの値の平均値を評価値として求める平均値算出手段と、各検索手法に対して得られた評価値を所定の順序に並べるソート手段とを設けた情報検索手法評価装置を提案する。

【0012】また、請求項5では、請求項4記載の情報検索手法評価装置において、人間の判断の結果を検索すべきとした人数と、検索すべきないとした人数を計測した後に、予め定められた式により人間の判断の違いの大きさを計算する計算手段と、人間の判断にノイズを与えるために、検索すべきとした人数と、検索すべきでないとした人数を比較する比較手段と、該比較手段により人数を比較した結果、多数の側の人数の値から予め定められた定数值を減じ、小数の側の人数の値に予め定められた定数值を加える加減算手段と、該加減算手段によって再設定された検索すべきとした人数、及び検索すべきでないとした人数から、予め定められた式に従って再度人間の判断の違いの大きさを計算する再計算手段と、全ての検索条件と検索データの組み合わせに対して、前記再計算手段による人間の判断とノイズの載った判断との違いの大きさの計算を行った結果の平均値を評価値とし

て求める第2の平均値算出手段と、前記平均値算出手段によって得られた評価値と前記第2の平均値算出手段によって得られた評価値とから予め定められた式により、評価値の最小値、最大値を計算する最小最大値算出手段と、評価値、最小値、最大値について、小さい順に並べることで各検索手法のそれぞれの順位を計算する順位算出手段と、評価値、最小値、最大値の3つの値それぞれについての各検索手法の順位の値を予め定められた式により一つの値とする順位決定手段とを設けた情報検索手法評価装置を提案する。

【0013】さらに、請求項6では、請求項4又は5記載の情報検索手法評価装置において、一つの検索対象データと一つの検索対象データとの間で、検索すべきとした人数と、検索すべきでないとした人数を比較する比較手段と、該比較手段の比較結果が同数であったならば予め定められた値を発生する定数発生手段を設けた情報検索手法評価装置を提案する。

【0014】

【作用】本発明の請求項1によれば、検索対象データとデータベース中の任意の一つのデータが、少なくとも一人以上の人間によって検索対象データに合致すると判定された外部評価値の個数と、合致しないと判定された外部評価値の個数がそれぞれ計測され、この一つのデータに対する情報検索手法が判定した検索対象データに合致するか否かの判断が、与えられた外部評価値による検索条件に合致するか否かの判断の多数の側と同じであるならば、予め与えられた定数値がデータ評価値とされ、小数の側と同じであるならば、多数の側と小数の側の外部評価値の個数から、予め与えられた式によって得られた値がデータ評価値とされ、与えられた全ての検索対象データとデータベース中のデータの組合せについて前記の一データ評価値が生成される。さらに、全ての一データ評価値の平均値が計算され、これを情報検索手法の検索効率の評価値とすることにより検索効率が一つの指標を用いて表現され、少なくとも一つ以上の情報検索手法に対して計算した評価値を順序づけることにより、各種情報検索手法の優劣が判定される。

【0015】また、請求項2によれば、前記一データ評価値を計算する際に、予め定められた式に従って、外部評価値の小数の側の人数と多数の側の人数の値から現評価値が計算され、多数の側の人数から予め定められた人数を減じ、小数の側の人数に多数側から減じた値を加えて、予め定められた式に従い仮想評価値が計算され、該仮想評価値と現評価値との差がデータ評価値の揺れ度合いとされ、全ての検索条件とデータベース中のデータの組み合わせに対する揺れ度合いの平均値が計算され、さらに前記検索効率の値に揺れ度合いの値を増減した信頼区間付き検索効率が計算され、該信頼区間付き検索効率の最大値と最小値及び前記一データ評価値から予め与えられた式により各検索手法に対して一つの検索効率の

値が計算されて、複数の情報検索手法の検索効率の優劣が判断される。

【0016】また、請求項3によれば、前記一データ評価値の計算において、外部評価値で検索対象データに合致するに該当する人数と、合致しないに該当する人数とが比較され、これらが同数であった場合には、予め定められた値がデータ評価値とされて、各検索手法の検索効率が計算される。

【0017】また、請求項4によれば、計測手段により、入力手段によって入力された検索すべきか否かの少なくとも一人以上の人間の判断のデータから、該被検索対象データが検索されるべきか、されないべきかを判定するために、人間の判断した検索すべきか否かの判断について、それぞれの人数が計測され、決定手段により、検索すべきか否かの判定を行うための予め与えられた値と検索すべきか否かの人数比の測定値とが比較され、該被検索対象データが検索すべきデータであるか否かが決定される。次に、判断手段によって、入力された情報検索システムの検索すべきか否かの判断結果と、人間の判断から決定された検索すべきか否かの判断とが比較され、検索システムによって得られた判断結果が人間による判断と一致しているかが決定される。さらに、人間の判断と検索手法の判断が一致している場合には、定数値発生手段によって、予め定められた定数値が与えられ、人間と検索手法の判断が異なっている場合には、ペナルティ値発生手段により、人間の判断における検索すべきか否かの判断のそれぞれの人数の割合の値から、予め定められた式によりペナルティの値が計算される。例えば、判断が一致した場合の定数値が0とされ、一致しなかった場合のペナルティの値が0よりも大きい値とされる。これにより検索すべきか否かの人間の判断に検索手法の判断が完全に一致する場合には、一つ一つの検索条件と一つ一つの検索対象データに対して常に0が与えられ、判断が一致しないデータに対しては、0よりも大きい値が得られるので、平均値算出手段によってペナルティ値の平均値を計算することにより、完全に判断が一致する場合には検索手法の評価値は0となり、判断が一致しない場合が多い検索手法ほど大きなペナルティの値をとり、各検索手法の能力がそれぞれ一つの数値として数値化される。さらに、計算されたペナルティの値を、ソート手段によって、例えば小さい順に順序付けすることにより少なくとも一つ以上の検索手法の能力を比較することができる。

【0018】また、請求項5によれば、一つ一つの検索条件と一つ一つの検索対象データの間の検索すべきか否かの判断を行う際に、計算手段によって、検索すべきとした人数と検索すべきでないとした人数の割合から予め定められた式に従って現評価値、即ち人間の判断の違いの大きさが計算され、比較手段によって、検索すべきとした人数と、検索すべきでないとした人数とが比較さ

れ、該比較結果に基づいて加減算手段により、検索すべきとした人数と検索すべきでないとした人数の多数の側の人数から予め定められた人数を減じ、小数の側に予め定められた人数が加えられる。この結果変更された人数の割合から予め定められた式に従って再計算手段によって再度計算が行われ、仮想評価値、即ちノイズの載った人間の判断の違いの大きさが計算されて、入力された人間による判断の結果と判断にノイズが載った結果が生成される。さらに、予め定められた式によってこれら2つの判断の違いの大きさが測定され、人間の判断が少し異なったときに、ペナルティの値にどのくらいの影響を与えるのかが計算される。この値は、十分に多くの人数によって、検索すべきか否かの判断が行われた場合には、極小数の人数が多数の側から小数の側へ移ったとしても、人数の割合に与える影響は大きく、計算されるペナルティへの影響の値も大きい。よってノイズが与えられた人間の判断と入力された人間の判断の違いの大きさを計算することで、どの程度信頼をおけるかが数値化できる。検索効率の評価値と、信頼の程度を示す値から、最小最大値算出手段により評価値の最小値、最大値が計算され、信頼度が低い場合には幅広い評価値の範囲、信頼度が高い場合には狭い評価値の幅の範囲データとして検索効率を表すことができる。さらに、最小値同士及び最大値同士、評価値同士で、効率の値の順位付けを行う順位算出手段と、予め定められた式により各値における順番から一つの値として検索効率の順番を計算する順位決定手段により、例えば信頼度の低い評価値の場合には、最大値が非常に大きくなる可能性が高いので、最大値、最小値、評価値の各値に対する検索手法の順序のうち、最大値の小さな検索手法がよい検索効率を示していると判定される。この結果、人間による判断において判定する人数が異なる場合でも、一つ以上の検索手法の善し悪しを自動的に判定できる。これにより、情報検索手法による結果と人間の判断が異なっていた場合に、検索すべきとした人の割合と、検索すべきではないとした人の割合からペナルティの値を計算するが、このペナルティの値は統計的に見れば、判断を行った人数によって信頼性が変化するので、前記請求項4記載の装置では、必ず同数の人間によって検索すべきか否かの判断を行わなければならぬが、請求項5記載の装置によればこれに限られることがない。

【0019】また、請求項6によれば、多数小数を判断する際に、比較手段によって、一つの検索対象データと一つの検索対象データとの間で、検索すべきとした人数と、検索すべきでないとした人数とが比較され、該比較結果に基づいて、これらの人数が同数である場合には、定数発生手段により、予め定められた定数が発生されることで、多数、小数がない場合でも評価値が計算される。これにより、人間による判断が検索すべきとした人数と、検索すべきではないとした人数が同数であった

場合にも、検索すべきデータなのか検索すべきでないデータなどを、判断することができる。

【0020】

【実施例】以下、図面に基づいて本発明の一実施例を説明する。図1は本発明の第1の実施例における情報検索手法評価装置を示す構成図であり、該装置はCPUを主体としたコンピュータ及びプログラムから構成されている。図において、1は外部評価値・検索手法一致判定モジュール、2は外部評価値集計モジュール、3は関係判断モジュール、4は定数値発生モジュール、5はペナルティ計算モジュール、6は平均ペナルティ計算モジュール、7は評価値バッファ、8はソートモジュールである。尚、ここでは各モジュールは、プログラム（ソフトウェア）の処理単位として構成されているが、ハードウェアによる構成も可能である。

【0021】外部評価値・検索手法一致判定モジュール1は、検索手法による検索すべきか否かの判断と人間にによって与えられた判断とを比較する。

【0022】外部評価値集計モジュール2は、少なくとも一人以上によって行われた人間の判断の結果を入力して、検索すべきとした人の数、検索すべきでないとした人の数を計測する。

【0023】関係判断モジュール3は、検索すべきとした人の数と、検索すべきでないとした人の数の割合から、検索すべきデータか否かを決定する。

【0024】定数値発生モジュール4は、外部評価値・検索手法一致判定モジュール1における判定により、検索手法による判断と人間の判断とが一致している場合に、予め定められた例数値を発生させる。

【0025】ペナルティ計算モジュール5は、検索手法による判断と人間の判断とが一致していない場合に、人間による検索すべきか否かの人数の割合から、検索手法が判断を誤ったことに対するペナルティ値を決定する。

【0026】平均ペナルティ計算モジュール6は、一つの検索手法に対する全ての検索条件と検索対象データの組み合わせに対して計算されたペナルティ値の平均をとる。

【0027】評価値バッファ7は、計算された平均値を評価値として検索手法と共に格納する。

【0028】ソートモジュール8は、評価値バッファ7中の検索手法を評価値の小さい順に並べ直す。

【0029】前述の構成よりなる本実施例によれば、検索手法の能力を測定する際には、予め装置外部から、各種検索手法によるデータベースの検索結果である情報検索手法検索結果表と、同じデータに対して人間が検索すべきか否かの判断を行った結果である外部評価地表が入力される。

【0030】これらの入力データの内、情報検索手法検索結果表は、図3に示すように、検索手法の手法の番号と、各検索条件とデータベース中の各データとの組を表

11

*識別子と、該検索手法が各識別子で表される組を検索すべきとしたか、否かを表す判断フラグから構成される。

【0031】また、外部評価値表は、図4に示すように、各検索条件とデータベース中の各データとの組を表す識別子と、各識別子で表現される検索条件とデータの組に対して、各人が検索すべきとしたか否かを示す判断フラグから構成されている。

【0032】第1の実施例における処理の過程としては、図5に示すように、まずステップ11として人間が判断した結果である外部評価値表のデータは、外部評価値集計モジュール4において検索すべきとした人数と検索すべきではないとした人数がそれぞれ数えられる。次に、ステップ12として関係判断モジュール3において、検索すべきとした人の人数と、検索すべきでないとした人の人数の割合から、検索すべきデータであるか、検索すべきデータではないかを判断し、判断の結果の値を検索すべきとした人の人数と検索すべきでないとした人の数と共に、外部評価値・検索手法一致判定モジュー*

$$\frac{\text{検索すべきとした人の数}}{\text{検索すべきとした人の数} + \text{検索すべきでないとした人の数}} > 0.7 \quad \cdots(3)$$

検索すべきデータであると判定することで実現できる。

【0035】次に一つ一つの検索条件とデータベースの中の一つ一つのデータに対して与えられている情報検索手法検索結果表のデータが、ステップ13として1外部評価値・検索手法一致判定モジュール1に送られ、モジュール内の評価値保存バッファに順次格納される。評価値保存バッファには、図6に示すように情報検索手法検索結果表から入力されたデータと、関係判断モジュール3から送られてきたデータが、対になって格納される。2つ以上の検索手法が評価される場合には、順次情報検索手法検索結果表からの入力データが置き換わり処理される。次に、ステップ14として外部評価値・検索手法一致判定モジュール1において、各検索条件・データ識別子に対する検索手法の結果と、関係判断モジュール3の結果が比較され、2つの値が等しい場合にはステップ15として定数値発生モジュール4に信号を送り、例えば0という値を生成して、平均ペナルティ計算モジュール6に送る。ステップ14における比較の結果、検索手法と関係判断モジュールによる判定の値とが異なっているときには、評価値保存バッファ中の検索すべきとした人数の値と検索すべきでないとした人数の値を、ペナルティ計算モジュール5に送る。ステップ16としてペナルティ計算モジュール5において、予め定められた式により、情報検索手法と関係判断モジュール3の値が異なっている場合の、ペナルティを計算する。

【0036】ステップ16における、ペナルティの計算の一実現法としては、予め定められた定数值を発生することで実現できる。例えば図6に示すように情報検索手法検索結果法の値と関係判断モジュールからの値が異なる場合、

10

12

*ル1に送られ、外部評価値・検索手法一致判定モジュール1に備わる外部評価値バッファ（図6に示す）に格納される。

【0033】前述したステップ12における検索すべきデータか否かの判断の方法として、検索すべきとした人の人数と、検索すべきでないとした人の人数とを比較して、多数の側であるとする事で実現できる。例えば図4のAAa009の場合、検索すべきであるとした人数は5人であり、検索すべきでないとした人数は2人であることから、検索すべきデータであると判定する。

【0034】ステップ12における検索すべきデータの他の判定方法としては、例えばAAa002の様に検索すべきとした人の数が4人であり、検査すべきでないとした人の数が3人であるように、人間による判断が割れてしまう場は無意味な情報を多く提供する可能性があるので、このようなデータは検索すべきでないとすると、例えば予め70%といったしきい値を定め、次の(3)式を満たす場合に、

【数2】

$$\frac{\text{検索すべきとした人の数}}{\text{検索すべきとした人の数} + \text{検索すべきでないとした人の数}} > 0.7 \quad \cdots(3)$$

* ているAAa001の様な場合、ステップ16により、例えば定数値として1を発生させ平均ペナルティ計算モジュール6へこの値を送る。

【0037】定数値発生モジュール4と、ペナルティ計算モジュール5によって得られた検索手法と人間の判断の一一致度を示す値は、ステップ17として平均値の計算を行い、検索手法全体の人間と検索手法の判断の一一致度を数値化した評価値を生成する。次にステップ18として、評価値が、検索手法に付けられている番号と共に評価値バッファ7に格納される。

【0038】与えられた全ての検索手法に対し、ステップ13～ステップ18間での処理が行われ、ステップ18-1において与えられた全ての検索手法に対しての処理が終わったことを判定すると、評価値バッファ7には検索手法の番号と評価値が格納され、ステップ19としてソートモジュール8において評価値の小さい順に並べ変える。ステップ19における並べ替えは、挿入法、ヒープソート、クイックソート等の既存の技術で容易に実現できる。また、ステップ20としてソートモジュール8で評価値の小さい順に並んだ検索手法の番号を出力することで、人間の判断に近い検索手法から順に選択することができる。

【0039】ステップ16におけるペナルティ計算の他の一実現法としては、図6におけるAAa100とAAb001を比較すると、AAa100はほとんどの人が検索すべきでないとしているにも拘らず、検索手法による結果が異なる判断をしているのに対し、AAb001では検索すべきとした人の数と検索すべきでないとした人の数がきわめて拮抗しているので、検索手法による判

40

13

断が検索すべきか否かのどちらであっても人間と異なった判断をしていると言えないため、人数比に応じてペナルティの値を変化させ、人間の意見と全く異なった場合には大きなペナルティを与え、人間の判断が曖昧な場合には小さなペナルティを与えることで、より詳細に人間の判断に近い検索手法から順に分類することもできる。これを実現するために、例えばペナルティ計算モジュール5を図7に示すような構成とする。

【0040】図7に示すペナルティ計算モジュール5は、入力された検索すべきとした人数と検索すべきでないとした人数とを一時的に格納する作業バッファ21と、作業バッファ21中の2つのデータの値を比較する比較モジュール22と、比較した結果、小さい方の値を分子、大きい方の値を分母として除算を行う除算モジュール23と、除された値を1から引いた値をペナルティ値とするペナルティ値発生モジュール24とから構成される。

【0041】このペナルティ計算モジュール5における処理の過程としては、図8に示すように、ステップ25として入力された検索すべきとした人数と検索すべきでないとした人数とを作業バッファ21に格納し、次にステップ26として作業バッファ21中の2つのデータの値を比較モジュール22において比較し、比較した結果小さい方の値を除算モジュール23の分子バッファ23bに、大きいほうの値を分母バッファ23aに格納し、ステップ27として分子バッファ23bと分母バッファ23aの値の除算を行い、最後にステップ28として、ペナルティ値発生モジュール24において除された値を1から引いた値をペナルティ値とする事で達成できる。

【0042】この実現法によれば、図6におけるAAa100の様に人間の判断がほぼ一致しているのにも拘らず、検索手法の判断が異なる場合には $1 - 1 \div 6 = 0.83$ となり、AAb001の様に人間の判断が割れている場合には、 $1 - 3 \div 4 = 0.25$ と小さなペナルティの値になる。この結果、人間の判断が割れている場合に多く人間の判断となる手法のペナルティ値は小さく、人間の判断が一致しているにも拘らず異なる判断を示す手

法は平均のペナルティ値が大きくなるので、両者を明確*

$$E = \frac{\log_{\frac{\text{検索すべきとした人数}}{\text{検索すべきでないとした人数}}}}{\log_{\frac{\text{検索すべきとした人数}}{\text{検索すべきでないとした人数}}}}$$

$$\frac{\log_{\frac{\text{検索すべきとした人数}}{\text{検索すべきでないとした人数}}}}{\log_{\frac{\text{検索すべきとした人数}}{\text{検索すべきでないとした人数}}}}$$

…(4)

ここで求めたエントロピーEを、ステップ38としてペナルティ値発生モジュール34において $(1-E)$ を計算してペナルティ値とする事でペナルティの計算を達成することができる。この実現法によれば、図6におけるAAa100の様に人間の判断がほぼ一致しているのにも拘らず、検索手法の判断が異なる場合には、 $1 - 0.99 = 0.01$ と小さなペナルティの値になる。この結果、人間の判断が割れている場合に多く人間の判断と異

*なる手法のペナルティ値は小さく、人間の判断が一致しているのにも拘らず異なる判断を示す手法は平均のペナルティ値が大きくなるので、両者を明確に分解することができる。

【0047】次に、本発明の第2の実施例を説明する。図11は、第2の実施例における情報検索手法評価装置の構成図であり、該装置はCPUを主体としたコンピュータ及びプログラムから構成されている。

14

*に分離することができる。

【0043】本発明におけるペナルティの考え方は、図6におけるAAa100の様に、ほとんどの人が検索すべきでないとしているにも拘らず、検索手法による結果が異なる判断をしていれば、検索手法が人間の判断に一致しないという情報の重要度は高く、AAb001の様に検索すべきとした人の数と検索すべきでないとした人の数が極めて拮抗しているのであれば、検索すべき否かは一概に決定できないので、人間の判断と一致しないという情報の重要性は小さいとみることができるということに基づいている。

【0044】この考え方は、情報のエントロピーの概念に類似する部分があるので、ステップ16の処理は、エントロピーの式を応用した手法によって、実現することも可能である。エントロピーの式を応用する場合の一実施例は、図9に示すような構成によって実現できる。図9に示すペナルティ計算モジュール5の一例は、入力された検索すべきとした人数と検索すべきでないとした人数とを一時的に格納する作業バッファ31と、作業バッファ31中の2つのデータの値から判断を行った総人數を計算する解答人數計算モジュール32と、検索すべきとした人数と検索すべきでないとした人数と解答人數計算モジュール32から得られた解答人數からエントロビ一値を計算するエントロピー計算モジュール33と、1からエントロビ一値を引いた値をペナルティ値とするペナルティ値発生モジュール34から構成される。

【0045】前述の構成よりなるペナルティ計算モジュール5における処理の過程としては、図10に示すように、ステップ35として入力された検索すべきとした人数と検索すべきでないとした人数とを作業バッファ31に格納し、次にステップ36として作業バッファ31中の2つのデータの値を解答人數計算モジュール32において足し合わせて解答した人數を計算し、次にエントロビ一計算モジュール33においてステップ37として以下に示す(4)式によりエントロピーの計算が行われる。

【0046】

【数3】

15

【0048】第2の実施例では図11に示すように、図1に示す第1の実施例とほぼ同様の構成をなしている。本実施例においては、外部から、各種検索手法によるデータベースの検索結果である情報検索手法検索結果表と、同じデータに対して人間が検索すべきか否かの判断を行った結果である外部評価値表を入力する。

【0049】入力されたデータから検索手法の能力を測定するために、第2の実施例の情報検索手法評価装置は、検索手法による検索すべきか否かの判断と人間によって与えられた判断とを比較する外部評価値・検索手法一致判定モジュール1と、少なくとも一人以上によって行われた人間の判断の結果を入力して、検索すべきとした人の数と、検索すべきでないとした人の数を計測する外部評価値集計モジュール2と、検索すべきとした人の数と、検索すべきでないとした人の数の割合から、検索すべきデータか否かを決定する関係判断モジュール3と、外部評価値・検索手法一致判定モジュール1における判定により、検索手法による判断と人間の判断とが一致している場合に、予め定められた定数值を発生させる定数值発生モジュール4と、検索手法による判断と人間の判断とが一致していない場合に、人間による検索すべきか否かの人数の割合から、検索手法が判断を誤ったことに対するペナルティ値を決定するペナルティ計算モジュール5と、一つの検索手法に対する全ての検索条件と検索対象データの組み合わせに対して計算されたペナルティ値の平均をとる平均ペナルティ計算モジュール6と、人間による検索すべきか否かの人数の割合から、人間の判断の情報量を計算する現評価値計算モジュール41と、人間の判断が多少変化していた場合の人間の判断の情報量を計算する仮想評価値計算モジュール42と、

計算された平均ペナルティ値と現評価値と仮想評価値を検索手法と組にして格納する中間バッファ43と、中間バッファ43中の検索手法を評価値と、現評価値と、仮想評価値とから順序付けする手法順位計算モジュール44の10個のモジュールから構成される。ここでも、第1の実施例と同様に、各モジュールは、プログラム（ソフトウェア）の処理単位として構成されているが、ハードウェアによる構成も可能である。

【0050】入力データである情報検索手法検索結果表と外部評価値表は、それぞれ前述した第1の実施例と同様であるので省略する。ただし、第1の実施例においては外部評価値表は、評価する全ての検索手法に対して一つしか存在しなかったが、本第2の実施例においては検索手法ごとに評価を行った人間の数が異なる外部評価値表を用いることもできる。

【0051】第2の実施例における処理の過程としては、図12に示すように、まずステップ51として人間が判断した結果である外部評価値表のデータは、外部評価値集計モジュール2において、検索すべきとした人数と検索すべきではないとした人数がそれぞれ数えられ、

10

20

30

40

50

16

関係判断モジュール3と現評価値計算モジュール41と、仮想評価値モジュール42に送られる。

【0052】次にステップ52として関係判断モジュール3において、検索すべきとした人の人数と、検索すべきでないとした人の人数の割合から、検索すべきデータであるか、検索すべきデータではないかを判断し、判断の結果の値を検索すべきとした人の数と検索すべきでないとした人の和と共に、外部評価値・検索手法一致判定モジュール1に送られ、図6に示す外部評価値バッファに格納される。

【0053】また、ステップ52における検索すべきデータか否かの判断の方法としては、第1の実施例と同様の方法がとれるので、その説明を省略する。

【0054】次に一つ一つの検索条件とデータベース中の一つ一つのデータに対して与えられている条件検索手法検索結果表のデータが、ステップ53として外部評価値・検索手法一致判定モジュール1に送られ、モジュール内の評価値保存バッファに順次格納される。評価値保存バッファは、図4に示すように情報検索手法検索結果表から入力されたデータと、関係判断モジュール3から送られてきたデータが、対になって格納される。2つ以上の検索手法が評価される場合には、順次検索手法検索結果表と外部評価値表からの入力データが置き換わり処理される。次にステップ54として外部評価値・検索手法一致判定モジュールにおいて、各検索条件・データ識別しに対する検索手法の結果と、関係判断モジュール3の結果が比較され、2つの値が等しい場合にはステップ55として定数值発生モジュール4に信号を送り、例えば0という値を生成し、平均ペナルティ計算モジュール6に送る。

【0055】ステップ54における比較の結果、検索手法と関係判断モジュール3による判定の値が異なっているときには、評価値保存バッファ中の検索すべきとした人数の値と検索すべきでないとした人数の値を、ペナルティ計算モジュール5に送る。ステップ56としてペナルティ計算モジュール5において、予め定められた式により、情報検索手法と関係判断モジュール3の値が異なっている場合の、ペナルティを計算する。

【0056】ステップ56における、ペナルティの計算の一実現法としては、例えば第1の実施例におけるエンタロピーを用いた手法を利用することができるので、ここでは説明を省略する。

【0057】定数值発生モジュール4とペナルティ計算モジュール5によって得られた検索手法と人間の判断の一致度を示す値は、ステップ57として平均値の計算を行い、検索手法全体に対する人間と検索手法の判断の一致度を数値化した評価値を生成する。

【0058】現評価値計算バッファ41においてはステップ58として、検索すべきとした人間の数と、検索すべきでないとした人間の数から、情報量の計算を行う。

17

情報量の計算は、例えば第1の実施例におけるエントロピーの計算と同様の手法で達成でき、容易に類推できるので説明を省略する。

【0059】仮想評価値計算バッファ42においてはステップ59として、検索すべきとした人間の数と、検索すべきではないとした人間の数の多数の側か予め定められた人数、例えば一人を減じ、小数の側に減じた人数を加えることで、人間の判断の結果にノイズを与えることができる。

【0060】次に、ステップ60としてノイズを与えた場合に対しての情報量の計算を行い、仮想評価値とする。仮想評価値の計算は、第1の実施例におけるエントロピーの計算と同様であるので、ここでの説明は省略する。現評価値計算モジュール41、及び仮想評価値計算モジュール42における処理は図11からも明らかなように、全ての場合について行われる。

【0061】次にステップ61として、評価値と現評価値と仮想評価値が、検索手法に付けられている番号と共に中間バッファ43に格納される。

【0062】与えられた全ての検索手法に対し、ステップ51～ステップ61間での処理が行われる。また、ステップ61-1において与えられた全ての検索手法に対する処理が終わったことが判定されると、中間バッファ43には検索手法の番号と評価値が格納され、ステップ62として手法順位計算モジュール44において評価値と現評価値、及び仮想評価値から、人間の判断に類似する順に検索手法の順序を決定する。さらに、ステップ63として、評価値の小さい順に検索手法番号を出力する。

【0063】手法順位計算モジュール44は、例えば図13に示す構成によって実現される。手法順位計算モジュール44は、中間バッファ43から送られてきた、評価値、現評価値、及び仮想評価値の3つの値から信頼区間を計算する信頼区間計算モジュール71と、評価値と計算された下限値及び上限値のそれぞれに対して検索手法の順位付けを行う順位付けモジュール72と、順位付けモジュール72で付けられた個々の検索手法に対する、下限、上限、評価値のそれぞれの順位から予め定められた式に従って、一つの順位値を生成する総合統一計算モジュール73と、計算された各検索手法の順位値を小さい順に並べて出力するソートモジュール74の4つのモジュールから構成される。

【0064】手法順位計算モジュール44における処理の過程としては、図14に示すように中間バッファ43から評価値と現評価値、及び仮想評価値を入力された信頼区間計算モジュール71において、ステップ81として例えば下限値、上限値を以下の(5)(6)式を用いて計算する。

【0065】

【数4】

18

$$\text{下限値} = \text{評価値} - \frac{|\text{現評価値} - \text{仮想評価値}|}{2} \quad \cdots(5)$$

$$\text{上限値} = \text{評価値} + \frac{|\text{現評価値} - \text{仮想評価値}|}{2} \quad \cdots(6)$$

計算された上限値、下限値と評価値は、ステップ82として順位付きモジュール72内のバッファに格納される。ステップ81、82を全ての検索手法の下限値及び上限値の計算が終了するまで繰り返す。この後、ステップ80において全ての検索手法について上限値及び下限値の計算が終了したか否かが判定され、計算が終了したら、ステップ83として順位付きモジュール72内の整列・順位決定部72aにおいて、各種法に対する下限値、上限値、及び評価値のそれぞれに対し小さい順に番号を割り当てる。この処理は、一般的なソーティングアルゴリズムにより容易に実現できる為、詳細な説明は省略する。

【0066】次にステップ84として、総合順位値計算モジュール73において、下限順位、評価順位、上限順位の値から総合順位を計算する。総合順位の計算方法は、例えば図13の総合順位値計算モジュール73内のバッファの図に示すように、各順位の総和によって求めることができる。

【0067】下限順位、上限順位のそれぞれの値は、下限順位は楽観的に見たときの検索効率を表し、上限値は悲観的に見たときの検索効率を表している。そこで総合順位値の他の一計算方法としては、検索手法に対して悲観的な見方をし、上限値の大きな手法を排除しようとするならば、例えば下限順位、上限順位、評価順位の総和をとる際に、上限順位の値をもう二度加える。すると、図13の場合では、手法Aは13、手法Bは8、手法Cは9、手法Dは20となり、手法Bの方が、順位値が小さくなる。

【0068】次にステップ85として、ソートモジュール74において総合順位値の小さい順に手法番号が出力される。ソートモジュール74における総合順位値の整列処理は、一般的なソーティングアルゴリズムにより容易に達成可能であるので、詳細な説明は省略する。

【0069】次に、本発明の第3の実施例を説明する。第3の実施例は前述した第1及び第2の実施例の双方に同一の構成を加えることで、実現可能であるので、ここでは第1の実施例を利用して説明を行う。

【0070】第3の実施例における構成図は、図15に示すように表すことができる。図15に示す構成においては、同数チェックモジュール9と同数時定数発生モジュール10が、第1の実施例の構成に付け加わっている。

【0071】第3の実施例における処理の過程は、図16に示すように、図5のステップ11とステップ12の間にステップ91として、同数チェックモジュール9に

19

おいて、検索すべきとした人の人数と検索すべきではないとした人の数を比較し、同数であるか否かを判定する。次に人数が同数の場合には、ステップ92として、同数チェックモジュール9から判定不可能として、関係判断モジュール3の結果の代わりに例えば-1と、検索すべきとした人数、検索すべきではないとして人数と共に、外部評価値・検索手法一致判定モジュール1に送られる。

【0072】検索すべきか否かの判定を行った結果の人数が同数でない場合には、ステップ12が実行される。ステップ12、13の処理は、第1の実施例と同様であるので、その説明は省略する。

【0073】次に、図5に示すステップ14の代わりにステップ93として、外部評価値・検索手法一致判定モジュール1において、各検索条件・データ識別子に対する検索手法の結果と、関係判断モジュール3の結果比較され、2つの値が等しい場合、若しくは、判定不可能の例えば-1がバッファ中の関係判断モジュール3による判定フラグに与えられていた場合には、ステップ15として同数時定数発生モジュール10に信号を送り、例えば0という値を生成して、平均ペナルティ計算モジュール6に送る。ステップ15以降の処理は、第1の実施例と同様であるのでその説明を省略する。

【0074】第2の実施例の手法に対して同様の機能を付加する場合についても、同様に図11における外部評価値集計モジュール2と、関係判断モジュール3の間に同数チェックモジュール9を加えることで達成できる。加えた場合の処理は、第1の実施例の構成に同数チェックモジュール9を加えた場合と同様であり、容易に類推できるのでその説明を省略する。

【0075】第3の実施例における、外部評価値・検索手法一致判定モジュール1におけるバッファ中のデータの一例は、図17に示すとおりである。

【0076】次に、本発明の能力を実験によって求めた*

検索手法	再現率	適合率	本発明の手法
AND検索	1.8%	50.0%	0.145
OR検索	100.0%	36.5%	0.318
最良検索	48.0%	96.5%	0.045
単語検索	55.3%	73.0%	0.076

表からも明らかなように、再現率と適合率ではOR検索がよいのか、最良検索がよいのか、単語検索がよいのかは判定できない。しかし、OR検索の場合には、大量の人間による判断では検索すべきでないデータが含まれていた。本発明の手法では、OR検索と、最良検索や、単語検索との間で明確に差が存在し、さらに人為的に高い検索効率を発生するような検索を行った最良検索がもっとも良いことを示している。この結果、本発明の手法では、良い検索手法から順に、最良検索、単語検索、AND検索、OR検索と自動的に検索手法の善し悪しを判断※50

20

*結果を以下に示す。

・実験条件：

実験に用いた検索対象データの数：12個（検索対象は、無作為に選ばれた新聞記事に付けられているキーワードと、新聞記事本文を構成する名詞、固有名詞、サ変名詞の集合の組からなる）

実際に用いた被検索データのサンプルの数：各条件に対して50個（無作為抽出）

比較した検索手法：4個

1. AND検索：検索対象中の全てのキーワードを持つ、被検索データを獲得する。

【0077】2. OR検索：検索対象中のいずれか一つ以上のキーワードを持つ、被検索データを獲得する。

【0078】3. 最良検索：検索対象中のいずれか一つ以上のキーワードを組み合わせて、評価用に適合率、再現率が共に高くなるように人為的に行った検索

4. 単語検索：検索対象中の本文を構成する名詞、固有名詞、サ変名詞の集合の、単語の一一致度の高いデータを獲得する。

【0079】被験者の数：9人

延べ評価値数： $12 \times 50 = 600$

・評価における計算法：第1乃至第3の実施例における、エントロピーの概念を応用した手法を用いた。この実験例では、被験者の数が奇数で且つ、人間による判断は同一なので、第2及び第3の実施例は、第1の実施例と同じとなる。なぜならば、第2の実施例の上限値と下限値の幅は、どの手法に対しても同じであり、また第3の実施例における検索すべきか否かの人間の判断が同数になることはない。

【0080】・結果：実験を行った結果を以下に示す。従来から用いられている、再現率、適合率と、本発明における評価値の関係は次の表の通りとなった。

【0081】

【表1】

※することができた。

【0082】

【発明の効果】以上説明したように本発明の請求項1記載の情報検索手法評価方法によれば、検索条件を入力してデータベース中から検索条件に合致するデータを獲得するデータベース検索手法において、少なくとも一人以上の人間が行った検索条件とデータベース中の各データ、もしくはデータベースから抽出されたサンプルデータに対して、検索条件に合致するか否かの判断のデータと、各検索手法によって得られた検索結果のデータとを

21

入力することで、少なくとも一つ以上の検索手法の検索効率の善し悪しを自動的に判定することができる。

【0083】また、請求項2によれば、上記の効果に加えて、人間の判断の結果は、多数の人間によって判断を行った場合と、1、2人の小数の人によって判断を行った場合には、統計的に見ると値の信頼性に差が生じるが、人間による評価における被験者の人が、検索手法によって異なっていたとしても、検索手法の優劣を判定することができる。

【0084】また、請求項3によれば、上記の効果に加えて、判断を行う人数が偶数個であり、検索すべきとする人数と検索すべきでないとする人数が同数となるても、一つの検索対象データと一つの検索対象データとの間で、検索すべきとした人数と、検索すべきでないとした人数を比較し、同数であったならば予め定められた値を発生することで、任意の人数による評価を行った場合でも検索手法の優劣を判定することができる。

【0085】また、請求項4記載の情報検索手法評価装置によれば、検索条件を入力してデータベース中から検索条件に合致するデータを獲得するデータベース検索手法において、少なくとも一人以上の人間が行った検索条件とデータベース中の各データ、もしくはデータベースから抽出されたサンプルデータに対して、検索条件に合致するか否かの判断のデータと、各検索手法によって得られた検索結果のデータとを入力することで、少なくとも一つ以上の検索手法の検索効率の善し悪しを自動的に判定することができる。

【0086】また、請求項5によれば、上記の効果に加えて、人間の判断の結果は、多数の人間によって判断を行った場合と、1、2人の小数の人によって判断を行った場合には、統計的に見ると値の信頼性に差が生じるが、人間による評価における被験者の人が、検索手法によって異なっていたとしても、検索手法の優劣を判定することができる。

【0087】また、請求項6によれば、上記の効果に加えて、判断を行う人数が偶数個であり、検索すべきとする人数と検索すべきでないとする人数が同数となるても、比較手段により、一つの検索対象データと一つの検索対象データとの間で、検索すべきとした人数と、検索すべきでないとした人数が比較され、同数であったならば定数発生手段により予め定められた値を発生することで、任意の人数による評価を行った場合でも検索手法の優劣を判定することができる。

【図面の簡単な説明】

【図1】本発明の第1の実施例における情報検索手法評価装置のシステム構成図

22

【図2】検索効率の評価法を示す概念図

【図3】第1の実施例における情報検索手法検索結果表のデータ構造とデータ例を示す図

【図4】第1の実施例における外部評価値表のデータ構造とデータ例を示す図

【図5】第1の実施例における情報検索手法評価方法の処理過程を示すフローチャート

【図6】第1の実施例における外部評価値バッファのデータ構造とデータ例を示す図

【図7】第1の実施例におけるペナルティ計算モジュールのシステム構成図

【図8】第1の実施例におけるペナルティ計算モジュールの処理の流れを示すフローチャート

【図9】ペナルティ計算モジュールの他の例を示すシステム構成図

【図10】他の例のペナルティ計算モジュールにおける処理流れを示すフローチャート

【図11】第2の実施例における情報検索手法評価装置のシステム構成図

【図12】第2の実施例における情報検索手法評価方法の処理の流れを示すフローチャート

【図13】第2の実施例における手法順位計算モジュールのシステム構成図

【図14】第2の実施例における手法順位計算モジュールの処理過程を示すフローチャート

【図15】第3の実施例における情報検索手法評価装置のシステム構成図

【図16】第3の実施例における情報検索手法評価方法の処理の過程を示すフローチャート

【図17】第3の実施例における外部評価値バッファのデータ構造とデータ例を示す図

【符号の説明】

- 1…外部評価値・検索手法一致判定モジュール、2…外部評価値集計モジュール、3…関係判断モジュール、4…定数発生モジュール、5…ペナルティ計算モジュール、6…平均ペナルティ計算モジュール、7…評価値バッファ、8…ソートモジュール、9…同数チェックモジュール、21…作業バッファ、22…比較モジュール、23…除算モジュール、24…ペナルティ値発生モジュール、31…作業バッファ、32…解答人数計算モジュール、33…エントロピー計算モジュール、34…ペナルティ値発生モジュール、41…現評価値計算モジュール、42…仮想評価値計算モジュール、43…中間バッファ、44…手法順位計算モジュール、71…信頼区間計算モジュール、72…順位付けモジュール、73…総合順位値計算モジュール、74…ソートモジュール

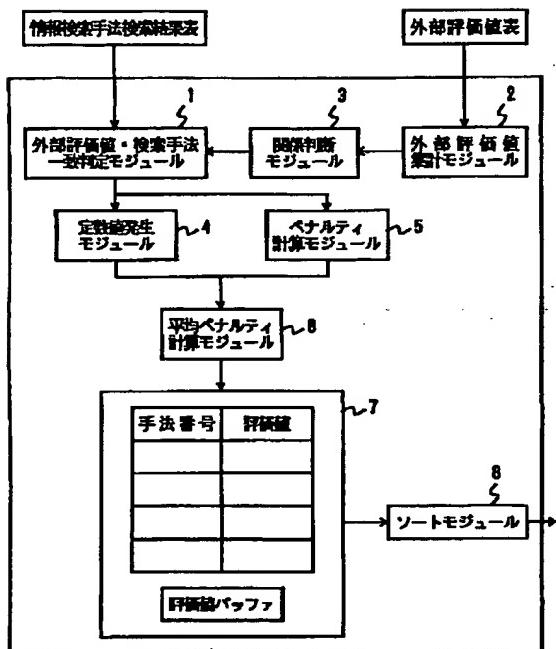
10

20

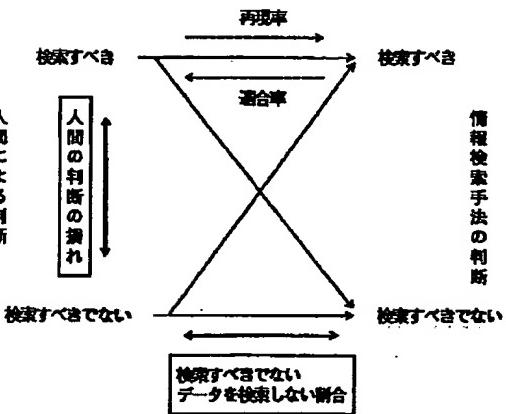
30

40

【図1】



【図2】



【図3】

検索手法名

1	検索手法番号
AND検索	
AAa001	0
AAa002	0
AAa003	1
AAa004	0
:	:
AAa099	0
AAa100	1
AAa101	0
AAb001	1
AAb002	0
AAb003	0
:	:

検索条件・データ選択子

半断フラグ (0:検索すべきでない, 1:検索すべき)

情報検索手法検索結果表

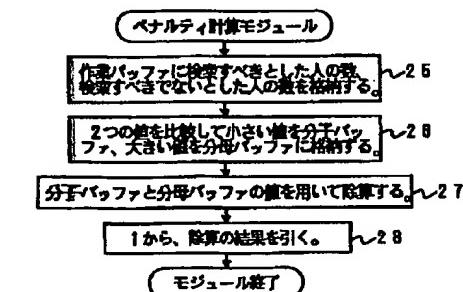
被験者 1の各 2の各 3の各 4の各 5の各 6の各 7の各

AAa001	1	1	0	1	0	1	1
AAa002	0	0	1	0	0	0	0
AAa003	1	1	1	1	0	1	1
AAa004	0	0	0	0	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
AAa099	1	1	1	0	0	1	1
AAa100	0	0	0	1	0	0	0
AAa101	0	0	0	0	1	0	0
AAb001	0	1	1	0	0	1	0
AAb002	0	0	1	1	1	0	1
AAb003	0	0	0	0	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

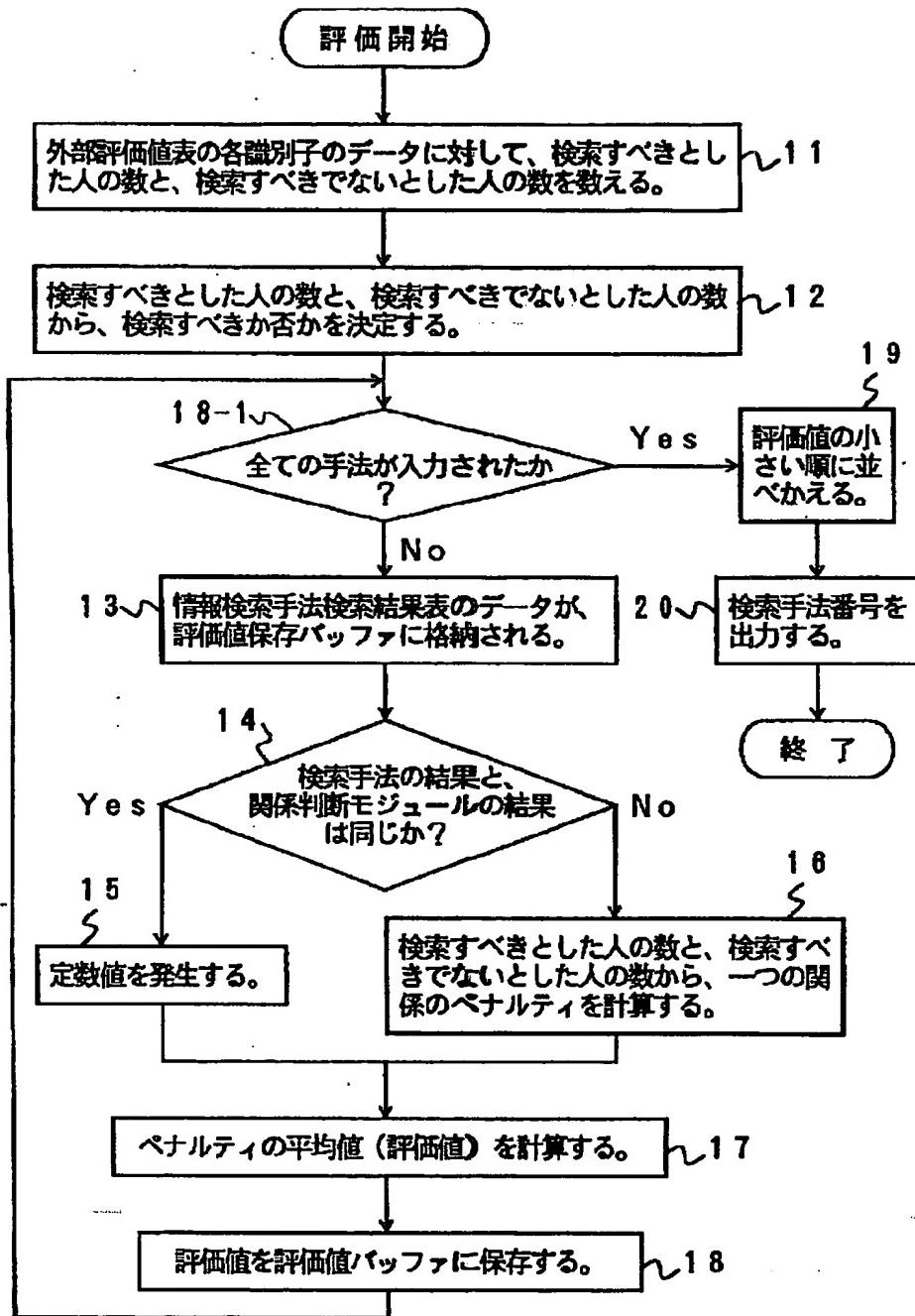
検索条件・データ選択子 半断フラグ (0:検索すべきでない, 1:検索すべき)

外部評価値表

【図4】



【図5】



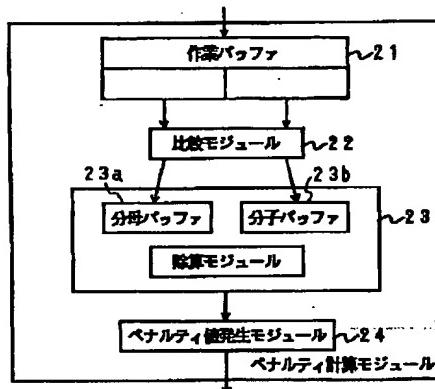
【図6】

AAa001	0	1	5	2
AAa002	0	0	1	6
AAa003	1	1	0	1
AAa004	0	1	5	7
:	:	:	:	:
AAa099	0	1	5	2
AAa100	1	0	1	6
AAa101	0	0	0	1
AAb001	1	0	3	4
AAb002	0	1	4	3
AAb003	0	0	0	7
:	:	:	:	:

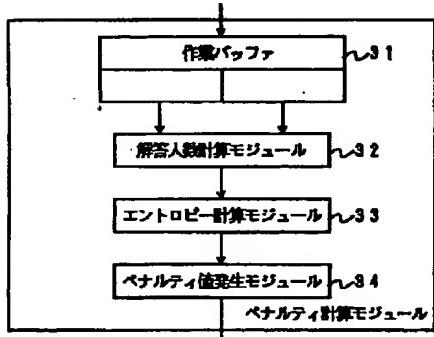
検索条件・情報検索手法検索結果
データ識別子表による判断フラグ 検索すべきと
検索すべきで
ないとした人数

回答評価モジュールによる判断フラグ
外部評価値バッファ

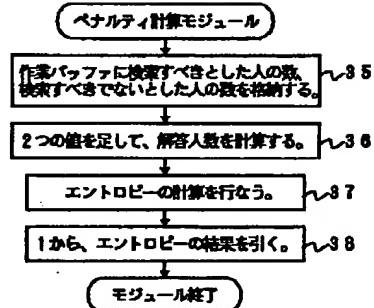
【図7】



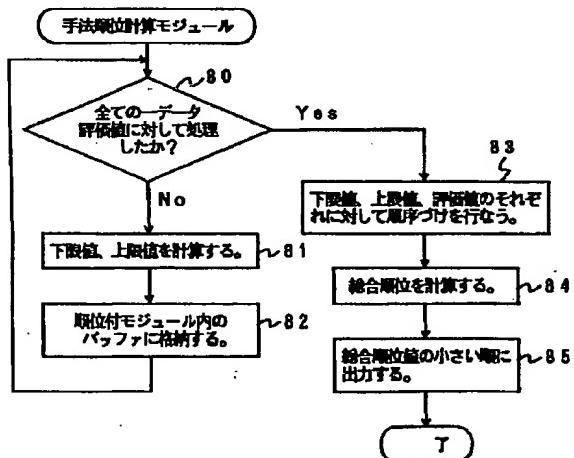
【図9】



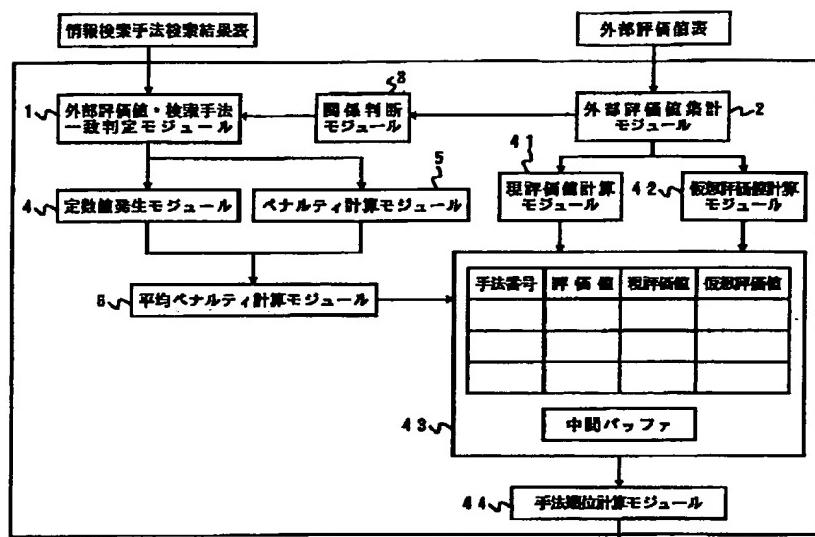
【図10】



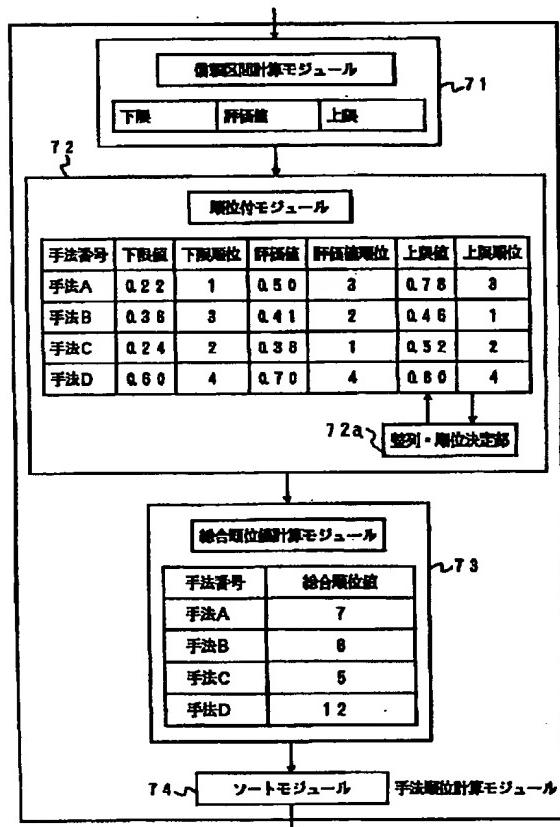
【図14】



【図11】



【図13】



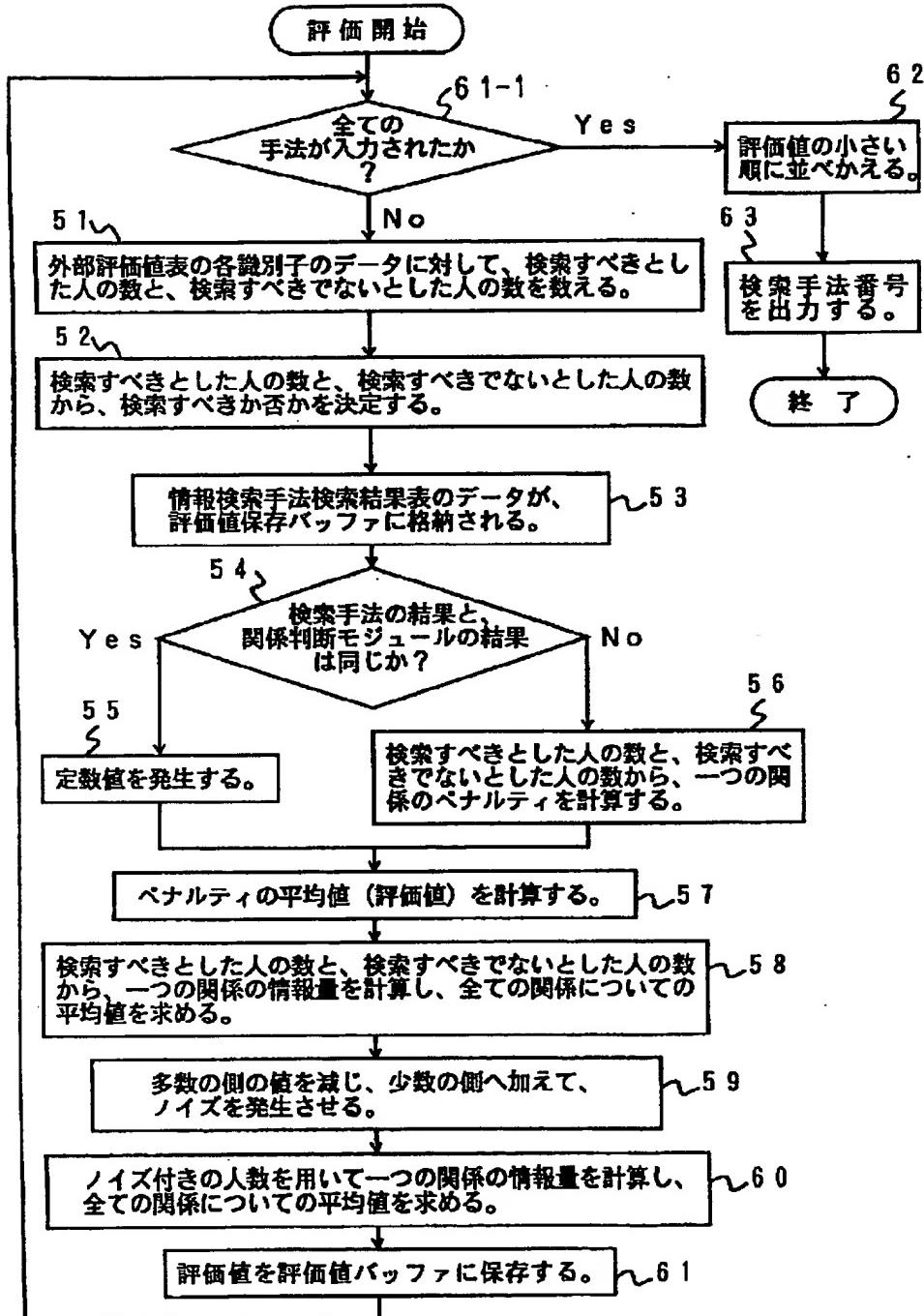
【図17】

AAa001	0	1	5	2
AAa002	0	0	1	0
AAa003	1	1	0	1
AAa004	0	0	0	7
:	:	:	:	:
AAa099	0	1	5	2
AAa100	1	0	1	0
AAa101	0	0	6	1
AAb001	1	0	3	0
AAb002	0	1	4	3
AAb003	0	0	0	7
:	:	:	:	:

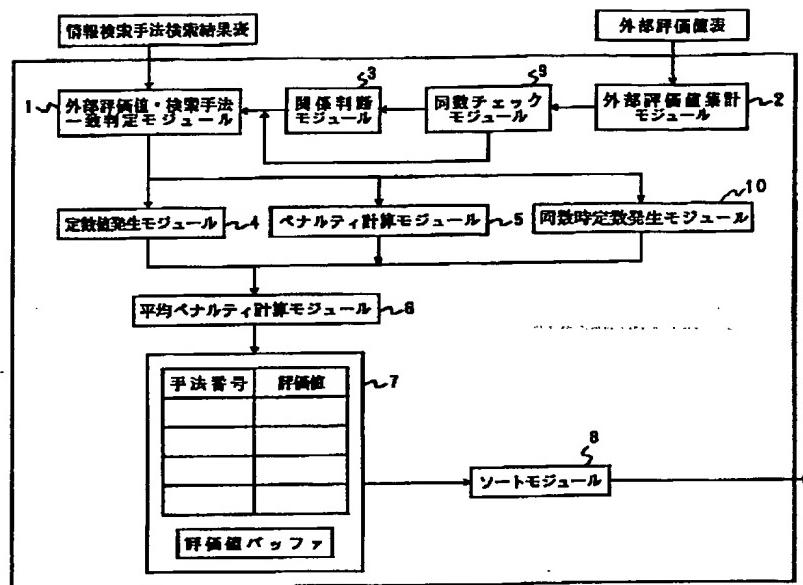
検索条件・データ属性による判断フラグ
情報検索手法検索結果
検索すべきとした人数
検索すべきでないとした人数
関係判断モジュールによる判断フラグ
外部評価値バッファ

This table displays a collection of data rows, likely representing search results or evaluation values. The columns correspond to the "AAa001" through "AAb003" entries shown in the main table above. Below the table, several descriptive labels are provided: "Search conditions · Data attributes judgment flag", "Information retrieval method search result", "Number of users to be searched", "Number of users not to be searched", "Relationship judgment module judgment flag", and "External evaluation value buffer".

【図12】



【図15】



【図16】

